

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AD1188  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-378 <GLA>  
A:Cross-references: GB:NC 003210; PIDN:CAC98986.1; PID:g16410311; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0908

Query Match 5.0%; Score 7; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 EMNKTIS 102  
|||||  
Db 214 EMNKTIS 220

RESULT 22  
AD1546  
membrane protein homolog lin0908 [imported] - *Listeria innocua* (strain Clp11262)  
C:Species: *Listeria innocua*  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AD1546  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.; Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunz, F.; Kurapk, G.; Madueno, E.; Maitournam, A.; Ma, O.K.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AD1546  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-378 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC96140.1; PID:g16413357; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lin0908

Query Match 5.0%; Score 7; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 EMNKTIS 102  
|||||  
Db 214 EMNKTIS 220

RESULT 23  
A70647  
probable PPE protein - *Mycobacterium tuberculosis* (strain H37RV)  
C:Species: *Mycobacterium tuberculosis*  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70647  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70647  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-409 <COL>  
A:Cross-references: GB:293867; GB:AL123456; NID:G3261695; PIDN:CAB06293.1; PID:el299813;  
A:Experimental source: strain H37RV  
C:Genetics:

Query Match 5.0%; Score 7; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 EMNKTIS 102  
|||||  
Db 214 EMNKTIS 220

A:Gene: PPE

Query Match 5.0%; Score 7; DB 2; Length 409;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 GASAIAA 63  
|||||  
Db 157 GASAIAA 163

RESULT 24  
A70932  
probable PPE protein - *Mycobacterium tuberculosis* (strain H37RV)  
C:Species: *Mycobacterium tuberculosis*  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70932  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70932  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-409 <COL>  
A:Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN:CAA17729.1; PID:el254f  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: PPE

Query Match 5.0%; Score 7; DB 2; Length 409;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 VLPQTGA 58  
|||||  
Db 342 VLPQTGA 348

RESULT 25  
ZZZRCL  
modulation protein nodC - *Rhizobium leguminosarum* plasmid pRL1J1  
C:Species: *Rhizobium leguminosarum*  
C>Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 21-Jul-2000  
C:Accession: A03486  
R:Roosen, L.; Johnston, A.W.B.; Downie, J.A.  
Nucleic Acids Res. 12, 9497-9508, 1984  
A:Title: DNA sequence of the *Rhizobium leguminosarum* modulation genes nodAB and C required for nodulation  
A:Reference number: A03482; MUID:85087952; PMID:6514582  
A:Accession: A03486  
A:Molecule type: DNA  
A:Residues: 1-424 <ROS>  
A:Cross-references: GB:X01650; NID:G46212; PIDN:CAA68619.1; PID:G46215  
C:Comment: This is one of the proteins, coded by nodulation genes, that are required for nodulation  
C:Genetics:  
A:Gene: nodC  
A:Genome: plasmid  
C:Superfamily: nodulation protein nodC  
C:Keywords: nodulation

Query Match 5.0%; Score 7; DB 1; Length 424;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIATV 65  
|||||  
Db 259 SAIATV 265

RESULT 26

F87108  
A:Molecule type: DNA  
A:Residues: 1-491 <STO>  
A:Cross-references: GB:AE004637; GB:AE004091; NID:g9948105; PIDN:AA05485.1; GSPDB:GN000000  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2097

Query Match 5.0%; Score 7; DB 2; Length 491;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GTGASAI 61  
|||||  
DB 183 GTGASAI 189

RESULT 29  
C70468  
phosphoribosylaminoimidazolecarboxamide formyltransferase - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 18-Jun-1999  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; C  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: C70468  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-506 <AQF>  
A:Cross-references: GB:AE000765; NID:g2984199; PIDN:AAC07734.1; PID:g2984204; GB:AE000000  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: purH  
C:Superfamily: purH bifunctional enzyme

Query Match 5.0%; Score 7; DB 2; Length 506;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 LRELSRK 93  
|||||  
DB 335 LRELSRK 341

RESULT 30  
D70861  
probable monooxygenase - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: D70861  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: D70861  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-524 <COL>  
A:Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDN:CAA16134.1; PID:el2377  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv3049c

Query Match 5.0%; Score 7; DB 2; Length 524;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GTGASAI 61

F87108  
probable amidase [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 30-Sep-2001  
C:Accession: F87108  
R:Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; MUID:21128732; PMID:11234002  
A:Accession: F87108  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-468 <STO>  
A:Cross-references: GB:AL450380; NID:gl3093388; PIDN:CAC30547.1; GSPDB:GN00147  
C:Genetics:  
A:Gene: MLI596  
C:Superfamily: indoleacetamide hydrolase

Query Match 5.0%; Score 7; DB 2; Length 468;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 SPELREL 90  
|||||  
DB 269 SPELREL 275

RESULT 27  
T28052  
hypothetical protein ZK858.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T28052  
R:White, S.  
submitted to the EMBL Data Library, September 1996  
A:Reference number: Z20462  
A:Accession: T28052  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-473 <WIL>  
A:Cross-references: EMBL:Z79759; PIDN:CAB02135.1; GSPDB:GN00019; CESP:ZK858.7  
A:Experimental source: clone ZK858  
C:Genetics:  
A:Gene: CESP:ZK858.7  
A:Map position: 1  
A:Introns: 23/3; 51/2; 177/3; 272/3; 338/3; 404/3; 456/3

Query Match 5.0%; Score 7; DB 2; Length 473;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 RLAMLRA 43  
|||||  
DB 219 RLAMLRA 225

RESULT 28  
F83383  
probable flavin-binding monooxygenase PA2097 [imported] - Pseudomonas aeruginosa (strain  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83383  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10584043  
A:Accession: F83383  
A:Status: preliminary

F87108  
A:Molecule type: DNA  
A:Residues: 1-491 <STO>  
A:Cross-references: GB:AE004637; GB:AE004091; NID:g9948105; PIDN:AA05485.1; GSPDB:GN000000  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2097

Query Match 5.0%; Score 7; DB 2; Length 491;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GTGASAI 61  
|||||  
DB 183 GTGASAI 189

RESULT 29  
C70468  
phosphoribosylaminoimidazolecarboxamide formyltransferase - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 18-Jun-1999  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; C  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: C70468  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-506 <AQF>  
A:Cross-references: GB:AE000765; NID:g2984199; PIDN:AAC07734.1; PID:g2984204; GB:AE000000  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: purH  
C:Superfamily: purH bifunctional enzyme

Query Match 5.0%; Score 7; DB 2; Length 506;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 LRELSRK 93  
|||||  
DB 335 LRELSRK 341

RESULT 30  
D70861  
probable monooxygenase - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: D70861  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: D70861  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-524 <COL>  
A:Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDN:CAA16134.1; PID:el2377  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv3049c

Query Match 5.0%; Score 7; DB 2; Length 524;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GTGASAI 61

F87108  
probable amidase [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 30-Sep-2001  
C:Accession: F87108  
R:Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; MUID:21128732; PMID:11234002  
A:Accession: F87108  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-468 <STO>  
A:Cross-references: GB:AL450380; NID:gl3093388; PIDN:CAC30547.1; GSPDB:GN00147  
C:Genetics:  
A:Gene: MLI596  
C:Superfamily: indoleacetamide hydrolase

Query Match 5.0%; Score 7; DB 2; Length 468;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 SPELREL 90  
|||||  
DB 269 SPELREL 275

RESULT 27  
T28052  
hypothetical protein ZK858.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T28052  
R:White, S.  
submitted to the EMBL Data Library, September 1996  
A:Reference number: Z20462  
A:Accession: T28052  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-473 <WIL>  
A:Cross-references: EMBL:Z79759; PIDN:CAB02135.1; GSPDB:GN00019; CESP:ZK858.7  
A:Experimental source: clone ZK858  
C:Genetics:  
A:Gene: CESP:ZK858.7  
A:Map position: 1  
A:Introns: 23/3; 51/2; 177/3; 272/3; 338/3; 404/3; 456/3

Query Match 5.0%; Score 7; DB 2; Length 473;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 RLAMLRA 43  
|||||  
DB 219 RLAMLRA 225

RESULT 28  
F83383  
probable flavin-binding monooxygenase PA2097 [imported] - Pseudomonas aeruginosa (strain  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83383  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10584043  
A:Accession: F83383  
A:Status: preliminary



```

Db      194 GTGASAI 200
|||||
RESULT 31
A:Accession: A83453
A:Title: Probable flavin-containing monooxygenase PA1538 [imported] - Pseudomonas aeruginosa (str)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83453
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:2043737; PMID:10984043
A:Accession: A83453
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <STO>
A:Cross-references: GB:AE004582; GB:AE004091; NID:9947492; PIDN:AG04927.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1538

Query Match      5.0%; Score 7; DB 2; Length 527;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      55 GTGASAI 61
|||||
Db      198 GTGASAI 204

RESULT 32
A:Accession: T40151
A:Title: histidine-tRNA ligase precursor, mitochondrial - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 19-Jul-2002
C:Accession: T40151
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z21842
A:Accession: T40151
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-538 <WOO>
A:Cross-references: EMBL:AL022103; PIDN:CAAI7892.1; GSPDB:GN00067; SPDB:SPBC2G2.12
A:Experimental source: strain 972h-; cosmid c2G2
C:Genetics:
A:Gene: SPDB:SPBC2G2.12
A:Map position: 2
A:Genome: nuclear
C:Superfamily: human histidine-tRNA ligase; amino acid-tRNA ligase repeat homology; hist
C:Keywords: mitochondrial
F:15-65/Domain: amino acid-tRNA ligase repeat homology <ATL>
F:66-526/Domain: histidine-tRNA ligase homology <HTL>

Query Match      5.0%; Score 7; DB 1; Length 538;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      130 RSLKSQ 136
|||||
Db      20 RSLKSQ 26

RESULT 33
A:Accession: S46458
A:Title: transcription factor tbx2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S46458

```

---

```

R;Bollag, R.J.; Siegfried, Z.; Cebra-Thomas, J.A.; Garvey, N.; Davison, E.M.; Silver, I
Nature Genet. 7, 383-389, 1994
A:Title: An ancient family of embryonically expressed mouse genes sharing a conserved f
A:Reference number: S46458; MUID:95004605; PMID:7920656
A:Accession: S46458
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-701 <BOL>
A:Cross-references: GB:U15566; NID:9558875; PIDN:AAC52697.1; PID:9558876
C:Genetics:
A:Gene: Tbx2
C:Superfamily: mouse transcription factor tbx2; T-box homology
F:104-285/Domain: T-box homology <TBX>

Query Match      5.0%; Score 7; DB 1; Length 701;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      128 HLRLSKS 134
|||||
Db      78 HLRLSKS 84

RESULT 34
A:Accession: G01840
A:Title: T-box protein 2 - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 23-Jul-1999
C:Accession: G01840
R;Campbell, C.E.
submitted to the EMBL Data Library, May 1995
A:Reference number: G08602
A:Accession: G01840
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-702 <CAM>
A:Cross-references: EMBL:U28049; NID:G924927; PIDN:AAA73861.1; PID:G924928
C:Genetics:
A:Gene: GDB:TBX2
A:Cross-references: GDB:568496; OMIM:600747
A:Map position: 17q21-17q22
C:Superfamily: mouse transcription factor tbx2; T-box homology
F:104-285/Domain: T-box homology <TBX>

Query Match      5.0%; Score 7; DB 2; Length 702;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      128 HLRLSKS 134
|||||
Db      78 HLRLSKS 84

RESULT 35
A:Accession: AH2546
A:Title: hypothetical protein all7659 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC71201
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH2546
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguci
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ar
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2546
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-731 <KUR>
A:Cross-references: GB:AP003602; PIDN:BAF77302.1; PID:g17134744; GSPDB:GN00181
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all7659

```

A:Genome: plasmid

Query Match 5.0%; Score 7; DB 2; Length 731;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 NKTISOE 104  
|||||  
Db 434 NKTISOE 440

RESULT 36  
F91089  
PTS system transcription regulator enzyme I [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: F91089  
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A95629; MUID:21156231; PMID:11258796  
A:Accession: F91089  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-748 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA037109.1; PID:gl33631158; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 050952  
C:Genetics:  
A:Gene: ECs3686

Query Match 5.0%; Score 7; DB 2; Length 748;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AMLRALA 45  
|||||  
Db 646 AMLRALA 652

RESULT 37  
H85934  
PTS system transcription regulator enzyme I [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 27-Nov-2001  
C:Accession: H85934  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: H85934  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-748 <STO>  
A:Cross-references: GB:AE005174; NID:gl2517314; PIDN:AAG57940.1; GSPDB:GN00145; UWGP:Z41  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: ptsP

Query Match 5.0%; Score 7; DB 2; Length 748;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AMLRALA 45  
|||||  
Db 646 AMLRALA 652

RESULT 38  
F65065  
phosphotransferase system enzyme I (EC 2.7.3.9) - Escherichia coli (strain K-12)  
C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: F65065  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F65065  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-748 <BLAT>  
A:Cross-references: GB:AE000366; GB:U00096; NID:gl789185; PIDN:AAC75868.1; PID:gl78919  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: ptsP  
C:Superfamily: phosphotransferase system enzyme I homology  
C:Keywords: phosphotransferase  
F:171-727/Domain: phosphotransferase system enzyme I homology <PT1>

Query Match 5.0%; Score 7; DB 2; Length 748;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AMLRALA 45  
|||||  
Db 646 AMLRALA 652

RESULT 39  
T32844  
hypothetical protein K05F6.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32844  
R:Du, Z.; Goela, D.  
submitted to the EMBL Data Library, December 1997  
A:Description: The sequence of C. elegans cosmid K05F6.  
A:Reference number: Z21233  
A:Accession: T32844  
A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA  
A:Residues: 1-753 <DUZ>  
A:Cross-references: EMBL:AF040653; PIDN:AA395025.1; GSPDB:GN00020; CESP:K05F6.5  
A:Experimental source: strain Bristol N2; clone K05F6  
C:Genetics:  
A:Gene: CESP:K05F6.5  
A:Map position: 2  
A:Introns: 28/2; 297/2; 343/3; 409/2; 551/2; 596/3; 666/3

Query Match 5.0%; Score 7; DB 2; Length 753;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ASAIAT 64  
|||||  
Db 720 ASAIAT 726

RESULT 40  
S59623  
tropoelastin - sheep  
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C:Date: 23-Aug-1996 #sequence\_revision 13-Mar-1997 #text\_change 24-Oct-1997  
C:Accession: S59623; A24758  
R:Mauch, J.C.; Sandberg, L.B.; Roos, P.J.; Jimenez, F.; Christiano, A.M.; Deak, S.B.; I  
Matrix Biol. 14, 635-641, 1994  
A:Title: Extensive alternate exon usage at the 5' end of the sheep tropoelastin gene.  
A:Reference number: S59623  
A:Accession: S59623  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-770 <NAU>  
R:Yoon, K.; Davidson, J.M.; Boyd, C.; May, M.; LuValle, P.; Ornstein-Goldstein, N.; Sm:

Arch. Biochem. Biophys. 241, 684-691, 1985  
A:Title: Analysis of the 3' region of the sheep elastin gene.  
A:Reference number: A24758; MUID:85305763; PMID:383997  
A:Accession: A24758  
A:Molecule type: mRNA  
A:Residues: 655-669,671-716,732-770 <YOO>  
C:Superfamily: elastin  
C:Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine  
F:760-765/Disulfide bonds: #status predicted

Query Match	5.0%	Score 7;	DB 2;	Length 770;
Best Local Similarity	100.0%	Pred. No. 79;		
Matches	7;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

Qy	52	VLPGTGA 58
Db	166	VLPGTGA 172

Search completed: November 14, 2003, 10:52:29  
Job time : 38 secs

Q54792	mus musculus	Q95429	rattus norv	Q95429	caenorhabdi	Q9x82	canis fami	Q95499	mycobacteri	Q53249	mycobacteri	Q33113	mycobacteri	Q3Y543	homo sapien	Q9xky0	homo sapien	Q43437	haemophilus	Q53376	mycobacteri	Q64171	mesocricetu	Q9xgq3	mus musculus	Q57086	homo sapien	Q02664	oryctolagus	Q02664	gallus galli	P56450	mus musculus	P56450	mus musculus	P02645	oryctolagus	P19317	homo sapien	P19313	rattus norv	P43367	escherichia	P97463	mus musculus	P24945	leptospira	P57069	vibrio chol	Q08638	archaeoglob	Q9C196	pasteurella	Q44116	synecchoc	Q31214	chromatium	Q9pni3	campylobact	Q02662	methanobact	Q57109	methanobact	Q99866	staphylobact	Q96K89	homo sapien	P13729	drosophila	Q9pdk5	xyella ffa	P76114	escherichia	Q9xkt2	vibrio chol	Q98364	xanthomonas	P23863	escherichia	Q82803	salmonella	Q8xgc4	salmonella	Q8xk76	xanthomonas	Q8xgb3	yersinia pe	P77201	escherichia	P28641	pisum sativ	P34594	caenorhabdi	Q9cpd1	pasteurella	P79204	ovis aries	Q99g87	rhizobium l	Q02835	archaeoglob	P54015	methanococ	Q9xvt6	pyrococcus	Q9xvr6	pyrobaculum	P45336	bacillus ca	Q9x7c2	mycobacteri	P78794	schizosacch	Q43819	homo sapien	Q8e8d9	shewanella	Q05620	canis fami	Q05620	neisseria m	P57014	neisseria m	P55044	neisseria m	Q53467	streptomyces	Q81275	arabidopsis	Q9xv216	pyrococcus	P59314	pyrococcus	Q57840	pyrococcus	Q921k8	halicobacte	Q25956	halicobacte	Q46600	arabidopsis	Q9px12	methanosa	P18623	streptomyce
--------	--------------	--------	-------------	--------	-------------	-------	------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	--------------	--------	-------------	--------	-------------	--------	--------------	--------	--------------	--------	--------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	--------------	--------	------------	--------	-------------	--------	-------------	--------	-------------	--------	-----------	--------	------------	--------	-------------	--------	-------------	--------	-------------	--------	--------------	--------	-------------	--------	------------	--------	------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	------------	--------	------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	------------	--------	-------------	--------	-------------	--------	------------	--------	------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	-------------	--------	------------	--------	------------	--------	-------------	--------	-------------	--------	-------------	--------	--------------	--------	-------------	---------	------------	--------	------------	--------	------------	--------	-------------	--------	-------------	--------	-------------	--------	-----------	--------	-------------

107	1	EFTS_MYCPE	Q8eug8 mycoplasma	180	6	4.3	404	1	ASSY_LISMO	Q8y5h2 listeria mo
108	1	TOXR_VIBPA	Q05938 vibrio para	181	6	4.3	404	1	YBR3_YEAST	P38083 saccharomyc
109	1	VG11_BPPH2	P04333 bacterioph	182	6	4.3	406	1	RL48_ARATH	Q9sf40 arabidopsia
110	1	TYSY_ENCCU	Q62584 encephalito	183	6	4.3	407	1	RL4A_ARATH	P49691 arabidopsia
111	1	DAPA_PROMA	P49423 prochloroco	184	6	4.3	413	1	NODC_RHISN	P50357 rhizobium s
112	1	EFTS_BAROU	Q8t66 bartonella	185	6	4.3	413	1	RF1M_YEAST	P30775 saccharomyc
113	1	EFTS_BAROU	Q9xcm5 bartonella	186	6	4.3	414	1	YLPD_CAEEL	P34374 caenorhabdi
114	1	HEM3_SALTY	Q816q2 salmonella	187	6	4.3	416	1	PXN1_XENLA	P49263 xenopus lae
115	1	A33_HUMAN	Q99795 homo sapien	188	6	4.3	417	1	PEDF_MOUSE	P97298 mus musculus
116	1	LIFA_ECOLI	P25845 escherichia	189	6	4.3	418	1	DCP1_TOBAC	P51845 nicotiana t
117	1	YC39_CYAPA	P48279 cyanophora	190	6	4.3	420	1	ISPG_BRUME	Q8y117 bruceella me
118	1	ASTE_ECOLI	P76215 escherichia	191	6	4.3	423	1	ARGD_KLULA	O14433 kluyveromyc
119	1	CTH1_YEAST	P47976 saccharomyc	192	6	4.3	424	1	CYAA_STIAU	P40137 stigmatella
120	1	CI30_HUMAN	Q9Y375 homo sapien	193	6	4.3	428	1	CLPX_XANAC	Q8bpi1 xanthomonas
121	1	SYFA_HELPU	Q9xkf9 helicobacte	194	6	4.3	428	1	CLPX_XANCP	Q8bpy5 xanthomonas
122	1	SYFA_HELPU	P56146 helicobacte	195	6	4.3	428	1	YB01_MYCPN	P75568 mycoplasma
123	1	DMPL_PSESP	P19730 pseudomonas	196	6	4.3	432	1	TIG_THETN	Q8xc26 thermoanaer
124	1	RECH_WIGBR	Q842w7 wisegsworth	197	6	4.3	433	1	HTR2_HALVA	P42258 haloarcula
125	1	YOO1_CAEEL	P34633 caenorhabdi	198	6	4.3	433	1	Y384_MYCCE	P47624 mycoplasma
126	1	MC4R_PIG	Q97504 sus scrofa	199	6	4.3	438	1	YC05_MYCPN	P75871 mycoplasma
127	1	MC4R_PIG	P70596 rattus norv	200	6	4.3	438	1	DNA4_THEME	P46798 thermotoga
128	1	SYFA_WIGBR	Q843b6 wiggleswort	201	6	4.3	446	1	DADR_DIDMA	P42288 didelphis m
129	1	MOAA_METAC	Q8tug2 methanosarc	202	6	4.3	446	1	DADR_HUMAN	P21728 homo sapien
130	1	MOAA_METMA	Q8px29 methanosarc	203	6	4.3	446	1	DADR_NACMU	Q77680 macaca mula
131	1	LPLA_ECOLI	P32099 escherichia	204	6	4.3	446	1	DADR_MOUSE	Q61616 mus musculus
132	1	FOSE_MOUSE	P13346 mus musculus	205	6	4.3	446	1	DADR_PIG	P50130 sus scrofa
133	1	SYFA_AQUAE	Q67087 aquifex ae	206	6	4.3	446	1	DADR_RAT	P18901 rattus norv
134	1	YA39_MYCPN	P75075 mycoplasma	207	6	4.3	449	1	IFR1_MOUSE	P13182 mus musculus
135	1	SYFA_CHLPP	Q926r6 chlamydia p	208	6	4.3	449	1	IFR1_RAT	P20695 rattus norv
136	1	NA37_HABIN	P44896 haemophilus	209	6	4.3	451	1	DADR_XENLA	P42288 xenopus lae
137	1	SYFA_CHLME	Q9p183 chlamydia m	210	6	4.3	451	1	IFR1_HUMAN	O00458 homo sapien
138	1	SYFA_MYCCE	P47436 mycoplasma	211	6	4.3	452	1	YEFF_ECOLI	P33016 escherichia
139	1	SYFA_CHLTR	O84843 chlamydia t	212	6	4.3	454	1	MUC_MESAU	Q06337 mesocricetu
140	1	UL21_HCNVA	P16846 human cytom	213	6	4.3	454	1	TRME_ECO57	Q8xb41 escherichia
141	1	GFDA_RHLO	P58142 rhizobium l	214	6	4.3	454	1	TRME_ECOLI	P25522 escherichia
142	1	SI4A_PIG	Q02745 s cmpn-n-ace	215	6	4.3	454	1	TRME_SALTI	Q822n8 salmonella
143	1	PYRC_NEIMA	Q9jvd6 neisseria m	216	6	4.3	454	1	TRME_SALTY	Q8xky3 salmonella
144	1	PYRC_NEIMB	Q9kod1 neisseria m	217	6	4.3	458	1	DBDR_XENLA	P42290 xenopus lae
145	1	Y240_MYCPN	P75442 mycoplasma	218	6	4.3	458	1	YNEI_CAEEL	P30640 caenorhabdi
146	1	FEN_METKA	Q8txu4 methanopyru	219	6	4.3	459	1	D1DR_FUGRU	P53452 fugu rubrip
147	1	AOX_ASPNG	O74180 aspergillus	220	6	4.3	459	1	IL7R_MOUSE	P16872 mus musculus
148	1	CCAP_METHY	Q26830 methanobact	221	6	4.3	460	1	MURD_ENTHR	Q07669 enterococcu
149	1	CCAP_ACEXY	P37697 acetobacter	222	6	4.3	460	1	NPL2_HUMAN	Q9ulw6 homo sapien
150	1	CUP5_GALME	Q24988 gallieria me	223	6	4.3	460	1	NPL2_MOUSE	P51860 mus musculus
151	1	Y4NL_RHISN	P55584 rhizobium s	224	6	4.3	463	1	DSDR_FUGRU	P53454 fugu rubrip
152	1	RUVB_STRCO	Q91391 streptomyce	225	6	4.3	463	1	DCDR_XENLA	P42291 xenopus lae
153	1	DHG_THEAC	P13203 thermoplasm	226	6	4.3	469	1	NRAM_IAPAR	P06819 influenza a
154	1	ALF_ECHMU	Q9gp32 echinococcu	227	6	4.3	471	1	VG20_BPP22	O01076 bacterioph
155	1	D1DR_CARAU	P35406 carassius a	228	6	4.3	474	1	VP61_NPVOP	O10270 orgyia pseu
156	1	RH23_SCHPO	O74803 schizosacch	229	6	4.3	480	1	SYFA_METJA	Q57911 methanococc
157	1	ID12_SULTO	Q96vw9 sulfolobus	230	6	4.3	481	1	PGKH_TOBAC	Q42961 nicotiana t
158	1	QPSV_APTME	O61303 apis mellif	231	6	4.3	481	1	SYFA_THEVO	Q979u4 thermoplasm
159	1	HNZ1_DROGU	Q24648 drosophila	232	6	4.3	485	1	NODC_BRAJA	P26024 bradyrhizob
160	1	YSAK_CAEEL	Q93454 caenorhabdi	233	6	4.3	486	1	QUIC_ACICA	Q10367 acinetobact
161	1	HSF5_ARATH	Q9s7u5 arabidopsia	234	6	4.3	486	1	YDBH_SCHPO	Q10367 schizosacch
162	1	DAPA_COILA	Q39535 coix lachry	235	6	4.3	488	1	UZIP_DROME	P10379 drosophila
163	1	NER2_MOUSE	Q9jmh3 mus musculus	236	6	4.3	490	1	C71P_ARATH	Q9stk8 arabidopsia
164	1	YO11_MOUSE	P11260 mus musculus	237	6	4.3	491	1	FIBB_HUMAN	P02675 homo sapien
165	1	DAPA_MAZE	P26259 zea mays (m	238	6	4.3	492	1	C136_MYCTU	P95099 mycobacteri
166	1	DLK_HUMAN	P80170 homo sapien	239	6	4.3	492	1	DPD2_HUMAN	Q9uh14 homo sapien
167	1	DLK_OREMO	P47800 oreochromis	240	6	4.3	495	1	GDF5_MOUSE	P43027 mus musculus
168	1	RL4_URECA	P49165 urechis cau	241	6	4.3	500	1	DPD2_RAT	Q9epb1 rattus norv
169	1	GCP_RICPR	Q9zeab rickettsia	242	6	4.3	501	1	GDF5_HUMAN	P43026 homo sapien
170	1	XYLR_HABIN	P45043 haemophilus	243	6	4.3	506	1	DPD2_MOUSE	Q9et22 mus musculus
171	1	HISZ_PSEAE	Q9uhm5 pseudomonas	244	6	4.3	509	1	GTR4_HUMAN	P14672 homo sapien
172	1	VATC_SCHPO	Q9hdw6 schizosacch	245	6	4.3	509	1	GTR4_MOUSE	P14142 mus musculus
173	1	EFTK_SCHPO	P87131 schizosacch	246	6	4.3	509	1	GTR4_RAT	P19357 rattus norv
174	1	RYST_STIAU	P42479 stigmatella	247	6	4.3	509	1	STK_HYDAT	P17713 hydra atten
175	1	NR21_ORILA	Q9y913 oryzias lat	248	6	4.3	512	1	THRC_ASHGO	Q00663 ashbya goss
176	1	ASSY_MEITH	Q27322 methanobact	249	6	4.3	514	1	PSBB_PROHO	P27200 prochloroth
177	1	BJAR_MOUSE	P25962 mus musculus	250	6	4.3	517	1	MURF_MYCLE	O69556 mycobacteri
178	1	YBJJ_ECOLI	P75810 escherichia	251	6	4.3	519	1	GALI_SCHOL	Q9hdu2 schizosacch
179	1	ASSY_LISIN	Q929s9 listeria in	252	6	4.3	523	1	E2BD_HUMAN	Q9hu10 homo sapien

253	6	4.3	523	1	E2BD_RABIT	P41111	oryctolagus	326	6	4.3	919	1	CAPP_COREF	O8rdl3	corynebacte
254	6	4.3	524	1	E2BD_MOUSE	Q61749	mus musculus	327	6	4.3	926	1	TAPT_SCHPO	Q14561	schizosacch
255	6	4.3	524	1	E2BD_RAT	Q63186	rattus norv	328	6	4.3	942	1	K6PF_SCHPO	Q42918	homo sapien
256	6	4.3	524	1	SVFA_METKA	Q8tym5	methanopyru	329	6	4.3	961	1	FGDI_HUMAN	P98174	homo sapien
257	6	4.3	525	1	RPS4_RHSN	P22881	rhizobium s	330	6	4.3	980	1	POLG_LIV	P22338	loupung ill
258	6	4.3	528	1	MDLC_PSPU	P20906	pseudomonas	331	6	4.3	988	1	M172_MOUSE	P97432	mus musculus
259	6	4.3	532	1	YABM_BACSU	P37555	bacillus su	332	6	4.3	990	1	YAA2_SCHPO	Q09796	schizosacch
260	6	4.3	532	1	HCP_CHLRE	Q8kbr4	chlorobium	333	6	4.3	998	1	VIOB_CHEVO	Q9s3v0	chromobacte
261	6	4.3	534	1	ILVD_THENA	Q9wz21	thermotoga	334	6	4.3	1004	1	SAL2_MOUSE	Q9qx96	mus musculus
262	6	4.3	556	1	CU48_SCHPO	P78750	schizosacch	335	6	4.3	1030	1	STK9_HUMAN	Q76039	homo sapien
263	6	4.3	581	1	LR15_HUMAN	Q8tf66	homo sapien	336	6	4.3	1055	1	XYNA_THENE	Q60042	thermotoga
264	6	4.3	589	1	DRYB_MOUSE	Q9z188	mus musculus	337	6	4.3	1060	1	EGS1_XENLA	P28025	xenopus lae
265	6	4.3	590	1	VPP_BPP2	P25479	bacterioph	338	6	4.3	1062	1	CARB_BACHD	Q9K9V9	bacillus ha
266	6	4.3	591	1	POXB_STRPN	Q54970	streptococc	339	6	4.3	1076	1	SPS3_SCHPO	Q13892	schizosacch
267	6	4.3	596	1	RGP1_DROME	Q9viw3	drosophila	340	6	4.3	1081	1	YF38_SCHPO	Q04933	craterostig
268	6	4.3	600	1	ABRA_PLARG	P23745	plasmodium	341	6	4.3	1091	1	NCA1_CHICK	P13590	gallus gall
269	6	4.3	604	1	ANYG_BHROT	P07683	rhizopus or	342	6	4.3	1145	1	POL_EIAVY	P03371	equine infe
270	6	4.3	605	1	WSC4_YEAST	P38739	saccharomyc	343	6	4.3	1146	1	POL_EIAVY	P11204	equine infe
271	6	4.3	607	1	BD10_NICSY	P46942	nicotiana s	344	6	4.3	1161	1	DANA_YEAST	P47179	saccharomyc
272	6	4.3	612	1	ADFI_CANAL	P46589	candida alb	345	6	4.3	1176	1	NIR_NEUCR	P38681	neurospora
273	6	4.3	617	1	ILVD_BUCDN	Q9rq56	buchnera ap	346	6	4.3	1176	1	SLAP_BACSH	P38537	bacillus sp
274	6	4.3	617	1	SYP_TREPA	Q83195	treponema p	347	6	4.3	1186	1	XPG_HUMAN	P28715	homo sapien
275	6	4.3	620	1	SVT_METJA	Q58597	methanococc	348	6	4.3	1186	1	XPG_HUMAN	P14629	xenopus lae
276	6	4.3	624	1	PMV_PASMU	Q9clv8	pasteurella	349	6	4.3	1196	1	DPOL_HSVBE	P28858	equine herp
277	6	4.3	627	1	ABPX_YEAST	Q08641	saccharomyc	350	6	4.3	1220	1	YMP3_CAEEL	P10947	caenorhabdi
278	6	4.3	629	1	DRB_HUMAN	Q9y463	homo sapien	351	6	4.3	1221	1	YMP3_CAEEL	P34472	caenorhabdi
279	6	4.3	630	1	INLB_LISWO	P25147	listeria mo	352	6	4.3	1222	1	YMP3_CAEEL	P51250	porphyra pu
280	6	4.3	630	1	MUC1_MOUSE	Q02496	mus musculus	353	6	4.3	1224	1	SLAP_BACST	P35825	bacillus st
281	6	4.3	630	1	YK01_SCHPO	O14115	schizosacch	354	6	4.3	1228	1	SLAP_BACST	P35825	bacillus st
282	6	4.3	638	1	CC45_SCHPO	O74113	schizosacch	355	6	4.3	1239	1	NNE3_MOUSE	Q01098	mus musculus
283	6	4.3	642	1	GATE_AERPE	Q9y9t6	aeropyrum p	356	6	4.3	1240	1	YQUB_CAEEL	Q09550	caenorhabdi
284	6	4.3	645	1	SVT_STAAM	Q99th9	staphylococ	357	6	4.3	1259	1	AUT2_HUMAN	Q8wx7	homo sapien
285	6	4.3	652	1	NAK1_SCHPO	O75011	schizosacch	358	6	4.3	1271	1	TPPS_XENLA	Q90286	xenopus lae
286	6	4.3	654	1	MCPC_BACSU	P54576	bacillus su	359	6	4.3	1271	1	PER1_HUMAN	O15534	homo sapien
287	6	4.3	656	1	YK79_MYCTU	Q10687	mycobacteri	360	6	4.3	1290	1	PER1_HUMAN	P17953	enterococcu
288	6	4.3	659	1	Y102_MYCLE	P53525	mycobacteri	361	6	4.3	1296	1	ASAI_ENTFA	P25455	drosophila
289	6	4.3	661	1	Y102_MYCTU	Q10897	mycobacteri	362	6	4.3	1312	1	PIP1_DROME	P12024	drosophila
290	6	4.3	680	1	OPDA_ECOLI	P27298	escherichia	363	6	4.3	1316	1	RPOC_MYCLE	P30761	mycobacteri
291	6	4.3	680	1	OPDA_SALTY	P27237	salmonella	364	6	4.3	1403	1	YGN1_YEAST	P53127	saccharomyc
292	6	4.3	683	1	APCE_SYNPE	P28035	synecococc	365	6	4.3	1500	1	GLF1_CANPA	P83509	canis fami
293	6	4.3	693	1	YQ01_SCHPO	Q14286	schizosacch	366	6	4.3	1504	1	SLIT_DROME	P24014	drosophila
294	6	4.3	694	1	TKT1_RHIME	P58333	rhizobium m	367	6	4.3	1508	1	GLF1_XANAC	P58938	xanthomonas
295	6	4.3	695	1	TKT2_RHIME	P56900	rhizobium m	368	6	4.3	1513	1	GLF1_XANAC	Q8nyr4	homo sapien
296	6	4.3	710	1	Y417_HUMAN	Q43301	homo sapien	369	6	4.3	1513	1	GLF1_XANAC	P81128	rattus norv
297	6	4.3	722	1	MAZ2_PSESM	P37330	escherichia	370	6	4.3	1526	1	MY52_SCHPO	Q9us16	schizosacch
298	6	4.3	725	1	YR65_CAEEL	Q87272	pseudomonas	371	6	4.3	1551	1	YQ12_CAEEL	Q09449	caenorhabdi
299	6	4.3	729	1	MASZ_AGR7S	Q92124	caenorhabdi	372	6	4.3	1589	1	CC25_YEAST	P04821	saccharomyc
300	6	4.3	731	1	MASZ_AGR7S	Q8uj85	agrobacteri	373	6	4.3	1589	1	PHP_DROME	P39769	drosophila
301	6	4.3	740	1	NLDL_HUMAN	Q8uqcl	homo sapien	374	6	4.3	1590	1	DC13_DROME	P18171	drosophila
302	6	4.3	741	1	BSQ2_DROME	P11929	drosophila	375	6	4.3	1693	1	SAS_DROME	Q04164	rattus norv
303	6	4.3	743	1	ABRA_PLARG	P26620	plasmodium	376	6	4.3	1711	1	PIPO_RAT	Q84612	rattus norv
304	6	4.3	748	1	PTIP_SALTY	P37178	salmonella	377	6	4.3	1839	1	CYAA_SACKL	P23466	saccharomyc
305	6	4.3	755	1	Y572_CHLPN	Q927Y1	chlamydia p	378	6	4.3	2204	1	RRPL_NDVB	P11205	newcastie d
306	6	4.3	760	1	SEPR_HUMAN	Q12884	homo sapien	379	6	4.3	2273	1	HPA1_YEAST	P32874	saccharomyc
307	6	4.3	761	1	SEPR_MOUSE	P97321	mus musculus	380	6	4.3	2524	1	NOTC_XENLA	P21783	xenopus lae
308	6	4.3	776	1	RTN1_HUMAN	Q16799	homo sapien	381	6	4.3	2556	1	NTC1_HUMAN	P46531	homo sapien
309	6	4.3	777	1	NTRY_AZOB	P45675	azospirillum	382	6	4.3	2670	1	IP3T_RAT	Q63269	rattus norv
310	6	4.3	779	1	PHK2_RHIME	Q822a4	rhizobium m	383	6	4.3	2671	1	IP3T_RAT	Q14573	homo sapien
311	6	4.3	783	1	HELS_HALNI	Q8hm66	halobacteri	384	6	4.3	2834	1	IF3R_DROME	P29993	drosophila
312	6	4.3	788	1	IR51_HCMVA	P09695	human cytom	385	6	4.3	2911	1	FN22_HUMAN	P35556	homo sapien
313	6	4.3	789	1	KR6P_HAECC	P27665	haemochrom	386	6	4.3	3172	1	ERY3_SACER	Q03133	saccharopol
314	6	4.3	800	1	T2D4_HUMAN	Q15542	homo sapien	387	6	4.3	3178	1	Y889_CAEEL	Q09620	caenorhabdi
315	6	4.3	827	1	ATCU_RHIME	Q9x5x3	rhizobium m	388	6	4.3	3301	1	CLR3_MOUSE	Q91210	mus musculus
316	6	4.3	829	1	E74A_DROME	P20105	drosophila	389	6	4.3	3312	1	CLR3_HUMAN	Q9nyg7	homo sapien
317	6	4.3	831	1	PLRL_MELEGRIS	Q10994	meleagris g	390	6	4.3	3313	1	CLR3_RAT	Q88278	rattus norv
318	6	4.3	846	1	IR51_HCMVA	P09715	human cytom	391	6	4.3	4563	1	AP3_HUMAN	P04114	homo sapien
319	6	4.3	858	1	ENV_HV2RO	P04577	human immun	392	6	4.3	4910	1	MON1_YEAST	Q12019	saccharomyc
320	6	4.3	878	1	CAPF_YERPE	Q8za84	yersinia pe	393	6	4.3	5120	1	PCLO_CHICK	Q29u36	gallus gall
321	6	4.3	878	1	MSH4_YEAST	P40965	saccharomyc	394	5	3.5	11	1	BPF3_BOTIN	P30423	bothrops in
322	6	4.3	879	1	E4L1_MOUSE	Q922h5	mus musculus	395	5	3.5	11	1	BPF4_BOTIN	P30424	bothrops in
323	6	4.3	881	1	E4L1_HUMAN	Q9t490	homo sapien	396	5	3.5	20	1	PL4_LUPLU	P83363	lupinus lut
324	6	4.3	881	1	HELI_HSVBE	P28934	equine herp	397	5	3.5	20	1	PL4_LUPLU	P83366	lupinus lut
325	6	4.3	913	1	E4L2_HUMAN	Q9h329	homo sapien	398	5	3.5	20	1	PL6_LUPLU	P83368	lupinus lut

399	1	DCUP_RHOSH	3.5	5	472	1	VKIL_LAMB	3.5	5	89	1	P31758 bacterioph
400	1	MT2_ERANA	3.5	5	473	1	Y008_TREPA	3.5	5	89	1	O83053 treponema p
401	1	LPID_ECOLI	3.5	5	474	1	Y007_SCHPO	3.5	5	89	1	O08678 schizosacch
402	1	PHNS_DESMU	3.5	5	475	1	Y2FA_ECOLI	3.5	5	89	1	P08339 escherichia
403	1	CHCD_ANTIPO	3.5	5	476	1	BAFL_DROME	3.5	5	90	1	Q9vlu0 drosophila
404	1	RL36_CORGL	3.5	5	477	1	NOLS_RHIME	3.5	5	90	1	Q52975 rhizobium m
405	1	Y04D_BPT4	3.5	5	478	1	RS20_RICPR	3.5	5	90	1	Q9zct1 rickettsia
406	1	YC17_CVACA	3.5	5	479	1	DBH_LACIA	3.5	5	91	1	Q9cti64 lactococcus
407	1	OPO2_OPICA	3.5	5	480	1	YFDM_ECOLI	3.5	5	91	1	P76509 escherichia
408	1	YC17_CVAPA	3.5	5	481	1	C553_BACPA	3.5	5	92	1	P82599 bacillus pa
409	1	STPI_BOVIN	3.5	5	482	1	RL37_EMENI	3.5	5	92	1	Q9C0t1 emericella
410	1	STPI_SHEEP	3.5	5	483	1	SH3N_HUMAN	3.5	5	93	1	Q9N299 homo sapien
411	1	ATPB_CORCN	3.5	5	484	1	YIAl_RHISP	3.5	5	93	1	P17983 rhizobium s
412	1	ATPB_MUSVO	3.5	5	485	1	NAG7_HUMAN	3.5	5	94	1	Q9Y6C7 homo sapien
413	1	BLYM_HUMAN	3.5	5	486	1	YGBF_ECOLI	3.5	5	94	1	P45956 escherichia
414	1	Y088_RICPR	3.5	5	487	1	HMVG_MOUSE	3.5	5	95	1	P17095 mus musculu
415	1	RZ0D_ECOLI	3.5	5	488	1	SR19_PYRAE	3.5	5	95	1	Q8ztb7 pyrobaculum
416	1	UCRY_SOLTU	3.5	5	489	1	VINI_BPT4	3.5	5	95	1	P31718 bacterioph
417	1	V238_FOWPV	3.5	5	490	1	YDHI_SCHPO	3.5	5	95	1	Q92346 schizosacch
418	1	LANH_STRMU	3.5	5	491	1	YNS9_XANAC	3.5	5	95	1	Q8pK18 xanthomonas
419	1	RPCX_SCHPO	3.5	5	492	1	TEN3_TENMO	3.5	5	96	1	Q27270 tenebrio m
420	1	IVBI_BUNFA	3.5	5	493	1	Y039_CHLNP	3.5	5	96	1	Q92922 chlamydia p
421	1	RL35_BUCBP	3.5	5	494	1	YS98_MYCLE	3.5	5	96	1	Q31024 mycobacteri
422	1	Y234_BRACH	3.5	5	495	1	D127_HUMAN	3.5	5	99	1	Q9hlma4 homo sapien
423	1	YN43_ARCFU	3.5	5	496	1	APA2_HUMAN	3.5	5	100	1	P02652 homo sapien
424	1	MP51_MYCAV	3.5	5	497	1	APA2_MACFA	3.5	5	100	1	P18656 macaca fasc
425	1	RS14_TOBAC	3.5	5	498	1	RS14_ANASP	3.5	5	100	1	O8Vg66 anabaena sp
426	1	XEN2_XENLA	3.5	5	499	1	SIS_LYNST	3.5	5	100	1	P42579 lymanaea sca
427	1	XEN3_XENLA	3.5	5	500	1	VP52_BPAPS	3.5	5	100	1	Q9t1p6 bacterioph
428	1	YSDA_ECOLI	3.5	5	501	1	Y837_NEIMB	3.5	5	100	1	Q91z25 neisseria m
429	1	ATPL_BACME	3.5	5	502	1	IF52_CHICK	3.5	5	101	1	Q09121 gallus gall
430	1	MARB_ECOLI	3.5	5	503	1	Y488_MYCPN	3.5	5	101	1	P75605 mycoplasma
431	1	Y055_NPVAC	3.5	5	504	1	YOF4_CAEEL	3.5	5	101	1	Q99269 caenorhadi
432	1	CSRA_BACSU	3.5	5	505	1	AMV1_HUMAN	3.5	5	103	1	O99417 homo sapien
433	1	RL29_STRCO	3.5	5	506	1	AMV1_MOUSE	3.5	5	103	1	O9egq3 mus musculu
434	1	YIAl_MYCTU	3.5	5	507	1	RL21_BORBU	3.5	5	103	1	O51719 borellia bu
435	1	YHNV_SALTY	3.5	5	508	1	Y446_NEIMA	3.5	5	103	1	Q9Jv14 neisseria m
436	1	APAZ_MACMU	3.5	5	509	1	H11_BOVIN	3.5	5	104	1	P02253 bos taurus
437	1	IF1C_CORMA	3.5	5	510	1	YS76_MYCTU	3.5	5	104	1	Q10802 mycobacteri
438	1	MT2B_ARATH	3.5	5	511	1	HMGI_CRIGR	3.5	5	106	1	Q9qxp3 cricetus
439	1	RL29_METKA	3.5	5	512	1	HMGI_HUMAN	3.5	5	106	1	P17096 homo sapien
440	1	RL29_MYCTU	3.5	5	513	1	Y4VC_RHISN	3.5	5	106	1	Q53211 rhizobium s
441	1	RUXX_ARCFU	3.5	5	514	1	I14K_TORMA	3.5	5	107	1	Q91499 torpedo mar
442	1	UL11_HSV6U	3.5	5	515	1	YMCD_ECOLI	3.5	5	107	1	P75885 escherichia
443	1	ATPB_ASFPM	3.5	5	516	1	HMGC_MOUSE	3.5	5	108	1	P52927 mus musculu
444	1	ATPB_STRDO	3.5	5	517	1	NIRD_SALTY	3.5	5	108	1	P40789 salmonella
445	1	NSGX_HUMAN	3.5	5	518	1	YIAl_MYCTU	3.5	5	108	1	Q50686 mycobacteri
446	1	RS18_UREPA	3.5	5	519	1	HMGC_HUMAN	3.5	5	109	1	P52926 homo sapien
447	1	YH65_SYNY3	3.5	5	520	1	RLAl_MAIZE	3.5	5	109	1	P52855 zea mays (m
448	1	Y222_METJA	3.5	5	521	1	YQCC_ECOLI	3.5	5	109	1	Q46919 escherichia
449	1	MT21_BRAJU	3.5	5	522	1	HDEA_ECOLI	3.5	5	110	1	P26604 escherichia
450	1	MT25_BRAJU	3.5	5	523	1	INS_PSAOB	3.5	5	110	1	Q62587 psammomya o
451	1	MT25_BRAJU	3.5	5	524	1	KV01_RABIT	3.5	5	110	1	P01882 cryptotlagus
452	1	RL4E_METJA	3.5	5	525	1	VG12_BPPFI	3.5	5	110	1	P25132 bacterioph
453	1	RL29_MYCLE	3.5	5	526	1	YC75_RALSO	3.5	5	110	1	Q8xqf2 raietonia s
454	1	TRUA_SYNP6	3.5	5	527	1	HM16_XENLA	3.5	5	111	1	P20912 xenopus lae
455	1	CXMB_CONMR	3.5	5	528	1	YE75_SHEON	3.5	5	111	1	Q8egw6 shewanella
456	1	Y056_NPVOP	3.5	5	529	1	PERX_PSEPU	3.5	5	112	1	P23103 pseudomonas
457	1	RR17_PORPU	3.5	5	530	1	H2B_TRYCR	3.5	5	112	1	P27795 trypanosoma
458	1	VG03_BPMI2	3.5	5	531	1	THIO_MYCSM	3.5	5	112	1	Q30974 mycobacteri
459	1	VG03_BPMI2	3.5	5	532	1	HYBF_KLEPN	3.5	5	113	1	Q9f0d4 klebsiella
460	1	R13B_STRAM	3.5	5	533	1	HYBF_MORMO	3.5	5	113	1	Q9zhri1 morganella
461	1	XEN3_XENLA	3.5	5	534	1	RBFA_OCEIH	3.5	5	113	1	Q8eqt9 oceanobacil
462	1	YS76_MYCLE	3.5	5	535	1	Y586_DEIRA	3.5	5	113	1	Q9rws9 deinococcus
463	1	RNFH_RHOCA	3.5	5	536	1	MYH8_MOUSE	3.5	5	113	1	Q9pmj7 campylobact
464	1	IRX3_HUMAN	3.5	5	537	1	THIO_MYCTU	3.5	5	115	1	P13542 mus musculu
465	1	RS20_ZYMMO	3.5	5	538	1	CHPE_ECOLI	3.5	5	115	1	P52229 mycobacteri
466	1	TAT_CAEVC	3.5	5	539	1	PFDB_ARCFU	3.5	5	116	1	P33647 escherichia
467	1	TAT_CAEVC	3.5	5	540	1	SP21_BACST	3.5	5	116	1	O29115 archaeglob
468	1	YCHL_LACIA	3.5	5	541	1	Y998_RHIME	3.5	5	116	1	Q52966 rhizobium st
469	1	REV_SIVAG	3.5	5	542	1	YA72_METJA	3.5	5	116	1	Q58472 methanococc
470	1		3.5	5	543							
471	1		3.5	5	544							

545	5	3.5	117	1	ARR1_ECOLI	P15905	esch	618	5	3.5	134	1	RS14_TORRU	Q9xek6	tortula
546	5	3.5	117	1	ARR2_ECOLI	P52144	esch	619	5	3.5	134	1	YR1_ECOLI	P10017	esch
547	5	3.5	117	1	GHRL_HUMAN	Q9ubq3	homo	620	5	3.5	135	1	YD39_SCHPO	Q10273	schiz
548	5	3.5	117	1	YGLJ_HSVB	P36344	simian	621	5	3.5	136	1	ACAC_PYRPU	P51796	pyrococ
549	5	3.5	118	1	YE26_RALS0	Q8xg99	rales	622	5	3.5	136	1	R141_YEAST	P06367	saccharo
550	5	3.5	119	1	YG58_METUA	Q59052	methanococ	623	5	3.5	137	1	IPPD_PIG	Q29277	sus scrofa
551	5	3.5	120	1	DAN3_YEAST	P38155	saccharomyc	624	5	3.5	137	1	NDK_STRCO	P50589	streptomyc
552	5	3.5	120	1	NLP7_GOSHI	Q24418	gossypium	625	5	3.5	137	1	NXT1_CAEEL	Q94757	caenorhabdi
553	5	3.5	120	1	NLP7_WAIZE	P19656	zea mays	626	5	3.5	137	1	R142_YEAST	P39516	saccharomyc
554	5	3.5	120	1	PAU1_YEAST	P38924	saccharomyc	627	5	3.5	137	1	RR9_CHLVU	P56358	chlorella v
555	5	3.5	120	1	PAU2_YEAST	P32612	saccharomyc	628	5	3.5	137	1	RS14_KLULA	P27069	kluyveromyc
556	5	3.5	120	1	PAU4_YEAST	P53427	saccharomyc	629	5	3.5	137	1	Y6C6_ECOLI	Q47719	esch
557	5	3.5	120	1	PAU6_YEAST	P52921	saccharomyc	630	5	3.5	137	1	YUKK_BACSU	P40761	bacillus su
558	5	3.5	120	1	SPI_HORSE	P81121	equus cabal	631	5	3.5	138	1	CXA4_PIG	Q29559	sus scrofa
559	5	3.5	120	1	YG65_YEAST	P53343	saccharomyc	632	5	3.5	138	1	RL16_MYCCE	P47404	mycoplasma
560	5	3.5	120	1	YG2F_YEAST	P53055	saccharomyc	633	5	3.5	138	1	Y29A_MYCCE	Q92b76	mycoplasma
561	5	3.5	120	1	YH56_YEAST	P38725	saccharomyc	634	5	3.5	139	1	ATPE_DICDH	P30159	dictyota di
562	5	3.5	120	1	YQXJ_BACSU	P24809	bacillus su	635	5	3.5	139	1	GOS9_ORYSA	P27349	oryza sativ
563	5	3.5	121	1	H1A_PLADU	P68894	platynereis	636	5	3.5	139	1	NUSA_HALMO	P15738	halococcus
564	5	3.5	121	1	RS11_UREPA	Q9pgn5	ureaplasma	637	5	3.5	139	1	PLAS_ANASO	O52830	anabaena sp
565	5	3.5	121	1	SR14_ARATH	O04421	arabidopsis	638	5	3.5	139	1	PLAS_ANASP	P46444	anabaena sp
566	5	3.5	121	1	VG38_HSV11	Q00144	ictaluriid h	639	5	3.5	139	1	PLAS_ANAVA	P00301	anabaena va
567	5	3.5	122	1	IN17_HUMAN	Q01444	ictaluriid h	640	5	3.5	139	1	RS14_SCHPO	P04129	schizosacch
568	5	3.5	122	1	RL18_THEMA	Q92ae3	thermotoga	641	5	3.5	139	1	Y201_RICPR	Q92dwa	rickettsia
569	5	3.5	122	1	YBDF_ECOLI	P39454	esch	642	5	3.5	139	1	YEL3_MYCPN	Q92d66	mycoplasma
570	5	3.5	123	1	PTHA_ECOLI	P05706	esch	643	5	3.5	140	1	Y337_MYCPN	P75297	mycoplasma
571	5	3.5	123	1	RK14_WAIZE	P05829	zea mays	644	5	3.5	141	1	HBAD_COTJA	P30892	cetrutrix co
572	5	3.5	123	1	RK14_ORYSA	P12137	oryza sativ	645	5	3.5	141	1	RL16_HELPY	Q92j60	helicobacte
573	5	3.5	123	1	RK14_TOBAC	P06382	nicotiana t	646	5	3.5	141	1	RL16_HELPY	P56041	helicobacte
574	5	3.5	123	1	RS12_BORBU	O51348	borrelia bu	647	5	3.5	142	1	Y535_PYRPU	Q8u3d3	pyrococcus
575	5	3.5	124	1	MAGS_HUMAN	P43359	homo sapien	648	5	3.5	143	1	PSB2_NICSV	Q41229	nicotiana s
576	5	3.5	124	1	PA21_BOTJA	P81243	bothrops ja	649	5	3.5	143	1	PTMA_STACA	P17876	staphylococ
577	5	3.5	124	1	RS12_TREPA	O83271	treponema p	650	5	3.5	143	1	RK13_GUITH	O46915	guillardi
578	5	3.5	124	1	VG19_BPM2	O64212	mycobacteri	651	5	3.5	143	1	SSRP_DEIRA	Q9rucl	deinococcus
579	5	3.5	124	1	YBF3_YEAST	P84190	saccharomyc	652	5	3.5	143	1	Y763_PYRHO	O58497	pyrococcus
580	5	3.5	125	1	SY02_RABIT	P28292	oryctolagus	653	5	3.5	143	1	RS14_TRYBB	P19800	trypanosoma
581	5	3.5	126	1	GP48_BPS17	O48402	bacterioph	654	5	3.5	144	1	Y4HN_RHISN	P55481	rhizobium s
582	5	3.5	126	1	RBFA_TREPA	O83860	treponema p	655	5	3.5	145	1	CAD3_PIG	O18926	sus scrofa
583	5	3.5	126	1	YF04_MYCPN	P75282	mycoplasma	656	5	3.5	145	1	DH11_GOSHI	P09442	gossypium h
584	5	3.5	126	1	YGM1_YEAST	P53130	saccharomyc	657	5	3.5	145	1	H2A1_WHEAT	P02275	tritium ae
585	5	3.5	127	1	RK12_ODOSI	P49550	odonella s	658	5	3.5	145	1	OM20_HUMAN	Q15388	homo sapien
586	5	3.5	127	1	SECE_ECOLI	P16320	esch	659	5	3.5	145	1	OM20_RAT	Q62760	rattus norv
587	5	3.5	127	1	SECE_SALTY	Q919k1	salmonella	660	5	3.5	145	1	PA21_LATLA	P19000	laticauda l
588	5	3.5	127	1	YFFN_ECOLI	P76545	esch	661	5	3.5	145	1	PA22_LATLA	Q8uu14	laticauda l
589	5	3.5	128	1	YH91_AQUAE	O67660	aquifex aeo	662	5	3.5	145	1	PA23_LATLA	Q8uu12	laticauda l
590	5	3.5	128	1	CRB2_HALNI	Q9hnm1	halobacteri	663	5	3.5	145	1	PA24_LATLA	Q8uu12	laticauda l
591	5	3.5	128	1	RS12_AQUAE	O70089	aquifex aeo	664	5	3.5	145	1	PA25_LATCO	Q8uu12	laticauda l
592	5	3.5	128	1	Y453_ARCFU	Q29796	archaeoglob	665	5	3.5	145	1	PA25_LATLA	Q8uu11	laticauda l
593	5	3.5	129	1	COAT_BPF2	P03611	bacterioph	666	5	3.5	145	1	PA25_LATLA	Q8uu11	laticauda l
594	5	3.5	129	1	COAT_BPM2	P03612	bacterioph	667	5	3.5	145	1	PA25_LATSE	Q91847	laticauda s
595	5	3.5	129	1	COAT_BPR17	P03613	bacterioph	668	5	3.5	145	1	PA25_LATSE	Q91846	laticauda s
596	5	3.5	129	1	COX8_HUMAN	P10606	homo sapien	669	5	3.5	145	1	PA25_LATSE	Q91844	laticauda s
597	5	3.5	129	1	DYLA_CHLRE	Q39591	chlamydomon	670	5	3.5	145	1	PA25_LATSE	Q91842	laticauda s
598	5	3.5	129	1	HMRK_RHILV	Q9x5v4	rhizobium l	671	5	3.5	145	1	PA25_LATSE	Q91837	laticauda s
599	5	3.5	129	1	MGSA_THETN	Q8rb66	thermoanaer	672	5	3.5	145	1	PA25_LATSE	O51450	borrelia bu
600	5	3.5	129	1	RK12_PORPU	P51339	porphyra pu	673	5	3.5	145	1	STIA_XENLA	Q90006	xenopus lae
601	5	3.5	129	1	RL3_YERPS	P11252	yersinia ps	674	5	3.5	145	1	YD00_PYRAB	Q9uz54	pyrococcus
602	5	3.5	129	1	RL3_YERPS	Q92j95	rickettsia	675	5	3.5	146	1	SPI7_RABIT	P36425	oryctolagus
603	5	3.5	129	1	RS12_RICCN	P41082	rickettsia	676	5	3.5	146	1	Y725_HAEIN	P44043	haemophilus
604	5	3.5	129	1	YBAZ_ECOLI	P75707	esch	677	5	3.5	146	1	Y273_CAEEL	Q19297	caenorhabdi
605	5	3.5	130	1	RL7_MYCTU	P37381	mycobacteri	678	5	3.5	147	1	IR13_HCMVA	P16811	human cytom
606	5	3.5	130	1	YGIW_ECOLI	P52083	esch	679	5	3.5	147	1	MEXR_PSEAE	P52003	psaedomonas
607	5	3.5	131	1	RS12_THETH	P17293	thermus the	680	5	3.5	147	1	NCZS_STRCZ	P01550	streptomyc
608	5	3.5	131	1	SPEN_THEAC	Q9hiV0	thermoplasm	681	5	3.5	148	1	RL9E_METJA	P54057	methanococ
609	5	3.5	132	1	RR12_CHLRE	P14149	chlamydomon	682	5	3.5	148	1	STNI_CHICK	P31395	gallus gall
610	5	3.5	132	1	RT09_SCHPO	O14006	schizosacch	683	5	3.5	148	1	STNI_HUMAN	P16949	homo sapien
611	5	3.5	132	1	SZ05_MOUSE	P50228	mus musculu	684	5	3.5	148	1	STNI_MOUSE	P54227	mus musculu
612	5	3.5	133	1	VPFI_LAMBD	Q3709	bacterioph	685	5	3.5	148	1	STNI_MOUSE	P13668	rattus norv
613	5	3.5	133	1	RNPA_COREF	Q8fsu8	corynebacte	686	5	3.5	148	1	TRBH_RHISN	P55405	rhizobium s
614	5	3.5	133	1	RNPA_CORGL	Q8n151	corynebacte	687	5	3.5	149	1	GLB1_MORMR	P21197	mordacia mo
615	5	3.5	133	1	RS9_CHLMU	Q9pkr2	chlamydia m	688	5	3.5	149	1	GLB1_PETWA	P05967	petromyzo
616	5	3.5	134	1	MGSA_CLOTS	Q9xcv1	clostridium	689	5	3.5	149	1	GLB3_MORMR	P21199	mordacia mo
617	5	3.5	134	1	RHOA_PLAFA	P11459	plasmodium	690	5	3.5	149	1	R141_WAIZE	P19950	zea mays



691	1	RL13	CHLPN	Q928t7	chlamydia p	764	5	3.5	158	1	ATPB	SCHGA	Q07233	schizaphis
692	1	RL9	VIBCH	Q9xuy9	vibrio chol	765	5	3.5	158	1	HLB2	ARATH	O24521	arabidopsis
693	1	RL9	VIBVU	Q8dcl3	vibrio vuln	766	5	3.5	158	1	KAB3	OLDAP	P58455	oldenlandia
694	1	RL9	XANAC	Q8gm12	xanthomonas	767	5	3.5	158	1	REG3	PYRAB	Q9v2d7	pyrococcus
695	1	RL9	XANCP	Q8pac0	xanthomonas	768	5	3.5	158	1	WMD	BACSU	P45910	bacillus su
696	1	RL9	XVIFA	Q9caf9	xylella fas	769	5	3.5	159	1	BV18	BETVE	P45431	betula verr
697	1	SP17	MOUSE	Q62252	mus musculus	770	5	3.5	159	1	BV18	BETVE	P43176	betula verr
698	1	YBAD	ECOLI	P25538	escherichia	771	5	3.5	159	1	BV18	BETVE	P43184	betula verr
699	1	YBAD	SALTY	Q8srd5	salmonella	772	5	3.5	159	1	BV18	BETVE	P43186	betula verr
700	1	YBAD	SHIFL	O51824	shigella fl	773	5	3.5	159	1	HLB2	GOSHI	Q93y92	gossypium h
701	1	AROQ	BUCAI	P57479	bughnera ap	774	5	3.5	159	1	MOAC	SHEON	Q8e942	shewanella
702	1	RL14	ARATH	Q9s1h0	arabidopsis	775	5	3.5	159	1	SSRP	RHILO	Q985b9	rhizobium l
703	1	RL14	ARATH	Q9cax6	arabidopsis	776	5	3.5	159	1	SSRP	RHILO	Q92r54	rhizobium m
704	1	RL14	MAIZE	P19951	zea mays (m	777	5	3.5	159	1	TRB6	AGRTU	P54916	agrobacteri
705	1	RL14	ARATH	P42036	arabidopsis	778	5	3.5	159	1	YC36	CYAPA	P48276	cyanophora
706	1	RL19	PRHNO	O59041	pyrococcus	779	5	3.5	160	1	MENG	DEIRA	Q95w10	deinococcus
707	1	RL9	VIBPA	Q87175	vibrio para	780	5	3.5	160	1	NIA	LOTTE	P39882	lotus tetra
708	1	RS14	LUPLU	O22584	lupinus lut	781	5	3.5	160	1	RS6	UREPA	Q9ppte6	ureaplasma
709	1	SSRP	THEMA	P56944	thermotoga	782	5	3.5	160	1	SSRP	AGRT5	Q8ugl2	agrobacteri
710	1	YEAL	HAEIN	P44110	haemophilus	783	5	3.5	161	1	HCRC	THAAR	O33818	thaueria aro
711	1	YP59	YEAST	Q02784	saccharomyc	784	5	3.5	161	1	OBP2	WANSE	P31419	manduca sex
712	1	DRPD	CRAPL	P22241	craterostig	785	5	3.5	161	1	PHAB	SYNEL	P50031	synecchococ
713	1	NRDI	MYCSA	Q9xc21	mycoplasma	786	5	3.5	161	1	TAT8	STRCO	Q9fbk8	streptomyce
714	1	RL19	SULSO	Q9ux89	sulfolobus	787	5	3.5	161	1	YA49	METJA	O58449	methanococ
715	1	RS13	AGABI	P78571	agarcicus bi	788	5	3.5	161	1	COAT	SHRV	P03581	sunh-hemp m
716	1	RS14	DROSOPH	P14130	drosophila	789	5	3.5	162	1	MAFG	HUMAN	O15525	homo sapien
717	1	RS14	HUMAN	P06366	homo sapien	790	5	3.5	162	1	MAFG	MOUSE	O54790	mus musculu
718	1	RS14	PODCA	Q08699	podocoryne	791	5	3.5	162	1	RRO	LVX	P27327	lilly virus
719	1	RS14	PROCL	P48855	procambatus	792	5	3.5	162	1	SNX3	YEAST	Q08826	saccharomyc
720	1	RS14	RAT	P13471	rattus norv	793	5	3.5	162	1	YB94	SCHPO	Q43003	schizosacch
721	1	SP17	HUMAN	O15506	homo sapien	794	5	3.5	163	1	RM23	YEAST	Q12487	saccharomyc
722	1	SP17	MACFA	P19021	macaca fasc	795	5	3.5	163	1	SP17	PAPHA	Q95230	papio hamad
723	1	MGSA	SALTY	Q15007	homo sapien	796	5	3.5	163	1	YB39	MYCPN	P75259	mycoplasma
724	1	MRA2	PASMU	Q8xer5	salmonella	797	5	3.5	164	1	IMI7	SCHPO	P87130	schizosacch
725	1	MRA2	SHVU	Q9cpb5	pasteurella	798	5	3.5	165	1	COAD	CHLTE	Q8kds9	chlorobium
726	1	MRA2	SHVU	Q9f9f9	shewanella	799	5	3.5	165	1	YSCH	YEREN	Q01249	yersinia en
727	1	PA2H	LATSE	Q8jfg2	laticauda s	800	5	3.5	165	1	YSCH	YERPE	Q00929	yersinia pe
728	1	RS14	CABEL	P48150	caenorhabdi	801	5	3.5	166	1	IPPI	RABIT	P01099	oryctolagus
729	1	RS14	NEUCR	P19115	neurospora	802	5	3.5	166	1	PV11	BPPRD	P27382	bacterioph
730	1	SOXR	SALTY	O56144	salmonella	803	5	3.5	167	1	BQCP	CHLPN	Q92901	chlamydia p
731	1	TCPS	VIBCH	P29484	vibrio chol	804	5	3.5	167	1	FTNB	ECOLI	P53091	escherichia
732	1	Y38A	MYCPN	P75213	mycoplasma	805	5	3.5	168	1	DYR	LACIA	Q59487	lactococcus
733	1	ECPC	ERICO	Q07564	eikenella c	806	5	3.5	168	1	PSA	DICDI	P12729	dictyostell
734	1	IFSA	HUMAN	P10159	homo sapien	807	5	3.5	168	1	YF24	MYCPN	P75254	mycoplasma
735	1	IFSA	RABIT	P10160	oryctolagus	808	5	3.5	170	1	CBP4	YEAST	P37267	saccharomyc
736	1	RECX	PSEAE	P37860	pseudomonas	809	5	3.5	170	1	PLAS	LYCEE	P17340	lycopersico
737	1	RISC	ACFCU	O28856	archaeoglob	810	5	3.5	170	1	PYRR	PSEAE	Q9x6w6	pseudomonas
738	1	RS14	CHLRE	P46295	chlamydomon	811	5	3.5	171	1	IM7A	HUMAN	Q99595	homo sapien
739	1	SP17	MACCU	O62770	macropus eu	812	5	3.5	171	1	IM7A	RAT	O35092	rattus norv
740	1	Y239	LISIN	Q92567	listeria in	813	5	3.5	172	1	CBR	DUNBA	P27516	dunaliella
741	1	YUGB	HAEIN	P44831	haemophilus	814	5	3.5	172	1	DEF	BORBU	O51092	borrelia bu
742	1	CYCP	CHRVU	P00154	chromatium	815	5	3.5	172	1	GSPH	PSEAE	Q00515	pseudomonas
743	1	PA26	LATCO	Q8uuh7	laticauda c	816	5	3.5	172	1	Y941	TREPA	O83911	treponema p
744	1	SOXR	ECOLI	P22538	escherichia	817	5	3.5	173	1	CF66	MOUSE	Q9dih6	mus musculu
745	1	SSRP	TREPA	O83214	treponema p	818	5	3.5	173	1	CHB2	LYMDI	P50604	lymantria d
746	1	YI23	ARCFU	O28452	archaeoglob	819	5	3.5	173	1	NADM	METTI	Q9uxn8	methanolobu
747	1	YN85	CABEL	P34598	caenorhabdi	820	5	3.5	173	1	RL10	SVNV3	P23350	synecchocyt
748	1	NUSB	RICPR	Q9ze01	rickettsia	821	5	3.5	174	1	ATP7	SCHPO	Q94390	schizosacch
749	1	PA23	LATCO	Q8uu10	laticauda c	822	5	3.5	174	1	EGD2	YEAST	P38879	saccharomyc
750	1	PA23	LATCO	Q8uuh9	laticauda c	823	5	3.5	174	1	HSCB	HAEIN	Q57006	haemophilus
751	1	PR13	PETCR	P19418	petroselinu	824	5	3.5	174	1	IPYR	THETH	P38576	thermus the
752	1	RL19	ARPER	Q9yif3	aeropyrum p	825	5	3.5	174	1	RS5	SYNEL	P59126	synecchococ
753	1	RL30	PYRAB	Q9v1v6	pyrococcus	826	5	3.5	174	1	Y110	MYCLE	Q9cd99	mycobacteri
754	1	RL30	PYRHO	O59440	pyrococcus	827	5	3.5	174	1	Y110	MYCTU	P72046	mycobacteri
755	1	HLB2	LYCES	Q941p9	lycopersico	828	5	3.5	174	1	Y110	MYCTU	O28784	archaeoglob
756	1	NUSB	RICCN	Q92165	rickettsia	829	5	3.5	174	1	YGR3	YEAST	P53102	saccharomyc
757	1	NSPN	MOUSE	O70300	mus musculus	830	5	3.5	175	1	PP18	STRCH	P77949	streptomyce
758	1	RS7	TREPA	O83272	treponema p	831	5	3.5	175	1	PYRE	HALNI	Q9hng2	halobacteri
759	1	YCO9	BRUME	O67324	aquifex aeo	832	5	3.5	175	1	SIVA	HUMAN	O15304	homo sapien
760	1	AROQ	SHVU	Q8ygu6	brucella me	833	5	3.5	175	1	Y240	METJA	Q57692	methanococ
761	1	IF51	SCHPO	P56289	schizosacch	834	5	3.5	176	1	MCH	METME	Q9xpw5	methylophil
762	1	IF52	SCHPO	Q9ust4	schizosacch	835	5	3.5	176	1	VG03	BPT4	P13331	bacterioph
763	1	RBFA	BIFLO	Q8g3y4	bifidobacte	836	5	3.5	176	1	YE45	METJA	Q58840	methanococ

837	1	DSBB_HAEIN	P44707 haemophilus	910	5	3.5	193	1	RLPB_ECOLI	P10101 escherichia
838	1	LB33_ARATH	Q91h88 arabidopsis	911	5	3.5	193	1	SYFB_ERWCH	P37984 erwinia chr
839	1	NREH_WOLSU	Q9sle6 wolinnella s	912	5	3.5	193	1	TRPF_STRMU	Q8dv4 streptococ
840	1	RL6_METH	Q26127 methanobact	913	5	3.5	193	1	VC07_ADE04	Q96831 human adeno
841	1	KCY_METJA	Q58071 methanococc	914	5	3.5	194	1	GRPE_BACHD	Q9kd73 bacillus ha
842	1	ADAB_BACSU	P19220 bacillus su	915	5	3.5	194	1	HI_SALIR	P02254 salmo brutt
843	1	CAS2_RAT	P02667 rattus norv	916	5	3.5	194	1	HIS7_SULTO	Q97021 sulfolobus
844	1	PURC_MYCLE	Q9rxb28 mycobacteri	917	5	3.5	194	1	IPPD_MOUSE	Q60829 mus musculu
845	1	RYVE_DEIRA	Q9rx75 deinococcus	918	5	3.5	194	1	RL6_SCHPO	P79071 schizosacch
846	1	SP17_MONDO	O62771 monodelphis	919	5	3.5	194	1	TRPG_HELPU	Q9zju6 helicobacte
847	1	RPF_CHLPN	Q9z7k6 chlamydia p	920	5	3.5	195	1	NUGM_MARPO	P34944 marchantia
848	1	SMCK_CRIGR	P41228 cricetus	921	5	3.5	196	1	AMIR_PSEAE	P10932 pseudomonas
849	1	Y521_PROMA	O51893 prochloroco	922	5	3.5	196	1	KTHY_ARCFU	O30175 archaeoglob
850	1	AG19_COCOPO	Q00398 coccidioid	923	5	3.5	196	1	NODA_RHILV	P04338 rhizobium l
851	1	RESA_BACSU	P35160 bacillus su	924	5	3.5	196	1	NUDH_HAEIN	Q57045 haemophilus
852	1	YG34_MYCPN	P75163 mycoplasma	925	5	3.5	196	1	R15E_HALNI	Q9hal2 halobacteri
853	1	YUAB_BACSU	P71014 bacillus su	926	5	3.5	196	1	R18A_HUMAN	Q9nvs62 homo sapien
854	1	HSLV_BORBU	Q57209 borrelia bu	927	5	3.5	196	1	RK3_ODOSI	P49569 odontella s
855	1	JDP_DROME	Q9tvp3 drosophila	928	5	3.5	196	1	VG53_BPT4	P16011 bacterioph
856	1	ORN_HAEIN	P45340 haemophilus	929	5	3.5	196	1	WBBJ_ECOLI	P16011 bacterioph
857	1	ACCI_DICDI	P14195 dictyosteli	930	5	3.5	196	1	YCHG_ECOLI	P37750 escherichia
858	1	CITC_ECOLI	Q8fk03 escherichia	931	5	3.5	196	1	YCHG_ECOLI	P37750 escherichia
859	1	CITC_ECOLI	P77563 escherichia	932	5	3.5	197	1	YV8B_YEAST	P40892 saccharomyc
860	1	Y6E2_ARCFU	O28810 archaeoglob	933	5	3.5	197	1	DOVR_RHOCA	P37740 rhodobacter
861	1	YGJUV_ECOLI	P42603 escherichia	934	5	3.5	197	1	DHNI_PEA	P28639 pisum sativ
862	1	Y3M9_MARPO	P38477 marchantia	935	5	3.5	197	1	HIS2_THEMEA	Q9x0c5 thermotoga
863	1	INB_RAT	P70499 rattus norv	936	5	3.5	197	1	LEXA_THEMEA	O86948 thermotoga
864	1	RBS5_ACECL	P16133 acetabulari	937	5	3.5	197	1	NRTN_HUMAN	Q97948 homo sapien
865	1	XY7_YEAST	O22205 saccharomyc	938	5	3.5	197	1	NUOC_RICPR	Q9zdh3 rickettsia
866	1	COM_DICDI	Q03380 dictyosteli	939	5	3.5	197	1	RECR_XANAC	O8p90 xanthomonas
867	1	RPSU_MYCTU	Q10679 mycobacteri	940	5	3.5	197	1	RECR_XANCP	O8p90 xanthomonas
868	1	RRF_CAMJE	Q9p1r0 campylobact	941	5	3.5	198	1	Y057_METKA	O8t282 methanopyru
869	1	RRF_CLOPE	Q46293 clostridium	942	5	3.5	198	1	SOC2_MOUSE	O35717 mus musculu
870	1	Y256_AERPE	Q9y7j2 aeropyrum p	943	5	3.5	198	1	SOC2_MOUSE	O85852 rattus norv
871	1	ATPD_RHORU	O5438 rhodospiril	944	5	3.5	198	1	YF93_YERPE	P58635 versinia pe
872	1	RFP_NEIMA	Q9juu2 neisseria m	945	5	3.5	199	1	CD69_HUMAN	P40325 saccharomyc
873	1	RFP_NEIMA	Q9jz98 neisseria m	946	5	3.5	199	1	RECO_HUMAN	O07108 homo sapien
874	1	SODC_NEIMA	P57005 neisseria m	947	5	3.5	199	1	RECO_HUMAN	P35343 homo sapien
875	1	SODC_NEIMA	O59623 neisseria m	948	5	3.5	199	1	RT14_CAEEL	P49391 caenorhabdi
876	1	TEBE_HALNI	Q9hns6 halobacteri	949	5	3.5	199	1	Y002_RHIME	Q92lf1 rhizobium m
877	1	YCEB_SALTY	P40822 salmonella	950	5	3.5	199	1	YA70_DEIRA	O8tgb8 monilia f
878	1	YJ41_PYRHO	O59604 pyrococcus	951	5	3.5	200	1	PHNE_PSEAE	Q9r1w2 campylobact
879	1	KIP2_HUMAN	O75838 homo sapien	952	5	3.5	200	1	REP2_ZYGFE	P09786 pseudomonas
880	1	KIP2_MOUSE	Q9z309 mus musculu	953	5	3.5	200	1	REP2_ZYGFE	P13742 zygocacchar
881	1	Y521_PYRAE	O8z444 pyrobaculum	954	5	3.5	200	1	RT26_RAT	Q9epj3 rattus norv
882	1	APT_NEIMA	Q9j195 neisseria m	955	5	3.5	200	1	YAG2_NEIMA	Q9j599 neisseria m
883	1	APT_NEIMA	Q9jy04 neisseria m	956	5	3.5	201	1	CUT1_MONFR	Q9julu4 neisseria m
884	1	APT_NEIMA	Q9kcc7 bacillus ha	957	5	3.5	201	1	CYP6_CAEEL	O8tgb8 monilia f
885	1	ISPZ_CAUCR	Q9a288 caulobacter	958	5	3.5	201	1	GCHI_STRCO	P52014 caenorhabdi
886	1	YAAH_ECOLI	P28695 escherichia	959	5	3.5	201	1	RECO_BOVIN	Q9x813 streptomyce
887	1	GCVR_ECOLI	P23483 escherichia	960	5	3.5	201	1	RECO_MOUSE	P21457 bos taurus
888	1	3MGH_DEIRA	Q9rsq0 deinococcus	961	5	3.5	201	1	RECR_YERPE	P34057 mus musculu
889	1	MAUE_METME	Q50231 methylophil	962	5	3.5	201	1	SDC2_HUMAN	Q8z97 yersinia pe
890	1	NUGM_ARATH	Q95748 arabidopsis	963	5	3.5	201	1	IPPD_BOVIN	P34741 homo sapien
891	1	NUGM_ORYSA	Q35322 oryza sativ	964	5	3.5	202	1	IPPD_BOVIN	P07516 bos taurus
892	1	NUGM_SOUTU	P80261 solanum tub	965	5	3.5	202	1	NLG3_MACMU	O8wmh2 macaca mula
893	1	VPSU_BPP4	P05460 bacterioph	966	5	3.5	202	1	PTH_TREPA	O83975 treponema p
894	1	YDO7_SCHPO	O13727 schizosacch	967	5	3.5	203	1	HIS5_VIBCH	Q9kx0 vibrio chol
895	1	CHB1_LYMDI	P43515 lymantria d	968	5	3.5	203	1	RECH_WYCTU	O6982 mycobacteri
896	1	Y645_METJA	O58061 methanococ	969	5	3.5	203	1	RL15_NEUCR	Q8x034 neurospora
897	1	RFP_BORBU	O51232 borrelia bu	970	5	3.5	203	1	RPOC_FISMU	P42074 fischerella
898	1	ILL8_PTG	O19073 sus scrofa	971	5	3.5	203	1	YNFI_HAEIN	P44102 haemophilus
899	1	NUGM_BETTR	Q33994 beta trigyn	972	5	3.5	203	1	YPT1_SCHPO	P11620 schizosacch
900	1	NUGM_BETTR	Q34011 beta webbii	973	5	3.5	204	1	IPPD_HUMAN	Q9ud71 homo sapien
901	1	NUGM_BETTR	P00300 serratia ma	974	5	3.5	204	1	KTHV_VACCV	P13410 vaccinia vi
902	1	TRPG_SERMA	P58997 escherichia	975	5	3.5	204	1	RECH_DEIRA	Q9rus2 deinococcus
903	1	BPPA_ECOLI	P33553 escherichia	976	5	3.5	204	1	SLI2_RALSO	Q9320 ralstonia s
904	1	BPPA_ECOLI	P33553 escherichia	977	5	3.5	204	1	VATD_AERPE	Q9y738 aeropyrum p
905	1	COAT_ACLSP	P27737 apple chlor	978	5	3.5	205	1	YTI6_CAEEL	Q10919 caenorhabdi
906	1	ILL8_BOVIN	Q9t473 bos taurus	979	5	3.5	205	1	ALAG_RAT	P02764 rattus norv
907	1	ILL8_CANFA	Q9xsr0 canis famli	980	5	3.5	205	1	COAT_TRVTC	P05072 tobacco rat
908	1	ILL8_HORSE	Q9xsg7 equus cabal	981	5	3.5	205	1	CYPB_YEAST	P23285 saccharomyc
909	1	PYRR_MYCTU	P71807 mycobacteri	982	5	3.5	205	1	CYSR_SYNY3	Q55854 synchococyt

P04892 bacterioph  
P46436 ascaris suu  
P33803 variola vir  
P71426 klebsiella  
Q9yb45 aetopyrum p  
Q9bvn8 homo sapien  
Q9hnt8 thermoplas  
Q63805 mus musculu  
P33970 halobacteri  
P22492 homo sapien  
P06350 oncorhynch  
Q8e688 shewanella  
Q8t290 pyrococcus  
Q8tjw9 mus musculu  
P21350 mus caroli  
P07361 mus musculu  
Q95833 homo sapien

ALIGNMENTS

RESULT 1  
OM22\_NEUCR STANDARD; PRT; 154 AA.  
AC Q07335;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Mitochondrial import receptor subunit TOM22 (Mitochondrial 22 kDa  
outer membrane protein) (MOM22 protein) (Translocase of outer membrane  
22 kDa subunit)  
GN TOM-22 OR TOM22 OR MOM22.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74A;  
RX MEDLINE=93351229; PubMed=8348615;  
RA Kiebler W., Keil P., Schneider H., van der Klei I.J., Pfanner N.,  
RA Neupert W.;  
RT "The mitochondrial receptor complex: a central role of MOM22 in  
mediating preprotein transfer from receptors to the general insertion  
pore."  
RL Cell 74:483-492 (1993).  
RN [2]  
RP FUNCTION.  
RX MEDLINE=96016192; PubMed=7556061;  
RA Mayer A., Margans F.E., Neupert W., Lill R.;  
RT "MOM22 is a receptor for mitochondrial targeting sequences and  
cooperates with MOM19."  
RL EMBO J. 14:4204-4211 (1995).

GIP complex proteins (By similarity).  
-1- SIMILARITY: BELONGS TO THE TOM22 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to a license@isb-sib.ch).  
-----  
CC EMBL; X71021; CAA50339.1; -.  
DR PIR; A40669; A40669.  
DR InterPro; IPR005683; Tom22.  
DR Pfam; PF04281; Tom22; 1.  
DR TIGRFAMs; TIGR00985; 3a0801s05tom22; 1.  
KW Receptor; Translocation; Transport; Protein transport; Outer membrane;  
KW Mitochondrion; Transmembrane. CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 1 84  
FT TRANSMEM 85 105  
FT DOMAIN 106 154  
FT DOMAIN 20 30  
FT ASP/GLU-RICH (ACIDIC).  
SQ SEQUENCE 154 AA; 16816 MW; C62EA2F3A7557439 CRC64;  
  
Query Match 5.0%; Score 7; DB 1; Length 154;  
Best Local Similarity 100.0%; Pred. No. 8,1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 5 STTNFV 11  
| | | | |  
Db 69 STTNFV 75  
  
RESULT 2  
U240\_DROME STANDARD; PRT; 203 AA.  
ID U240\_DROME  
AC Q9VH39;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Protein CG11722.  
DE CG11722 OR BCDNA:AT14909.  
GN Drosophila melanogaster (Fruit fly).  
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.V., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.B., Kodra C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamis I., Simpson M., Skupski M.P., Smith T.,  
 RA Sier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wagsman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhen M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Berkeley; TISSUE=Testis;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celnik S.E.;  
 RT "A *Drosophila* full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0240 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE003685; AAF54481.1; -;  
 DR EMBL; AY070793; AAL48415.1; -;  
 DR FlyBase; FBGN0037777; CG11722.  
 DR SEQUENCE 203 AA; 23705 MW; 325F385D98DC5F99 CRC64;  
 SQ  
 Query Match 5.0%; Score 7; DB 1; Length 203;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 29 LLRVKSS 35  
 DB 192 LLRVKSS 198  
 RESULT 3  
 ID MOTB TREPA STANDARD; PRT; 238 AA.  
 AC 007887;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chemotaxis motB protein (Motility protein B).  
 GN MOTB OR TP0724.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetes.  
 OX NCBI\_TaxID=160;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=96345631; PubMed=8755894;  
 RA Limberger R.J., Slivinski L.L., El-Afandi M.C.T., Dantuono L.A.;  
 RT "Organization, transcription, and expression of the 5' region of the  
 RT fla operon of *Treponema phagedenis* and *Treponema pallidum*.";  
 RL J. Bacteriol. 178:4628-4634(1996).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Nichols;  
 RC MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham D.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Artach P., Bowman C., Coston M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of *Treponema pallidum*, the syphilis  
 RT spirochete.";  
 RL Science 281:375-388(1998).  
 CC -1- FUNCTION: REQUIRED FOR THE ROTATION OF THE FLAGELLAR MOTOR. MIGHT  
 CC BE A LINKER THAT FASTENS THE TORQUE-GENERATING MACHINERY TO THE  
 CC CELL WALL (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.  
 CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U28219; AAB61254.1; -;  
 DR EMBL; AE001244; AAC65689.1; -;  
 DR FIR; H71291; H71291.  
 DR TIGR; TP0724; -;  
 DR InterPro: IPR006665; OmpA/MotB.  
 DR Pfam: PF00691; OmpA; 1.  
 DR ProDom; PD000930; OmpA/MotB; 1.  
 KW Chemotaxis; Flagella; Transmembrane; Inner membrane;  
 KW Flagellar rotation; Complete proteome.  
 FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 15 34 POTENTIAL.  
 FT DOMAIN 35 238 PERIPLASMIC (POTENTIAL).  
 SQ SEQUENCE 238 AA; 26050 MW; OAB5FCDBE9852E1F CRC64;  
 Query Match 5.0%; Score 7; DB 1; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 26 REALLRV 32  
 DB 138 REALLRV 144  
 RESULT 4  
 ID FPG\_MYCTU STANDARD; PRT; 289 AA.  
 AC Q10559;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) (FAPY-DNA  
 DE glycosylase).  
 GN MUTM OR PPG OR RV2924C OR MT2994 OR MTCY338.13C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=R37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Broech R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth I., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornaby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of *Mycobacterium tuberculosis* from the

RT complete genome sequence.";  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CDC 1551 / Oshkosh;  
 RC Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kelsoy J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Usterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS ENZYME MAY PLAY A SIGNIFICANT ROLE IN PROCESSES  
 CC LEADING TO RECOVERY FROM MUTAGENESIS AND/OR CELL DEATH BY  
 CC ALKYLATING AGENTS (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY. Hydrolysis of DNA containing ring-opened N(7)-  
 CC methylguanine residues, releasing 2,6-diamino-4-hydroxy-5-(N-  
 CC methyl)formamidopyrimidine.  
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PFG FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z74697; CA98987.1; -;  
 CC EMBL; AE007121; AAK47321.1; -;  
 CC PIR; D70748; D70748.  
 CC HSP; O50606; 1EE8.  
 CC TIGR; MT2994; -;  
 CC TubercuList; RV2924c; -;  
 CC HAMAP; MF\_00103; -;  
 CC InterPro; IPR000191; Fapy DNA glyco.  
 CC InterPro; IPR000214; Fapy DNA glyco.zn.  
 CC Pfam; PF01149; Fapy DNA glyco. 1.  
 CC ProDom; PD003680; Fapy DNA glyco. 1.  
 CC TIGRFAMS; TIGR00577; fpg; 1.  
 CC PROSITE; PS01242; PFG; 1  
 CC DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;  
 KW Complete proteome.  
 FT ZN FING 260 283 POTENTIAL.  
 SQ SEQUENCE 289 AA; 31950 MW; E00B94A70DC2904E CRC64;

Query Match 5.0%; Score 7; DB 1; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 DVMREAL 29  
 DB 219 DVMREAL 225

RESULT 5  
 ID\_NIFH\_CVAA5 STANDARD; PRT; 327 AA.  
 AC 007641;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nitrogenase iron protein (EC 1.18.6.1) (Nitrogenase component II)  
 DE (Nitrogenase Fe protein) (Nitrogenase reductase).  
 GN NIFH.  
 OS Cyanotheca (strain ATCC 51142).  
 OC Bacteria; Cyanobacteria; Chroococcales; Cyanotheca.  
 OX NCBI\_TaxID=43989;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20062360; PubMed=10594374;

RA Colon-Lopez M.S., Tang H.-Y., Tucker D.L., Sherman L.A.;  
 RT "Analysis of the nifHDK operon and structure of the NifH protein from  
 RT the unicellular, diazotrophic cyanobacterium, *Cyanothece* sp.  
 RL ATCC 51142.";  
 RL Biochim. Biophys. Acta 1473:363-375 (1999).  
 CC -1- FUNCTION: THE KEY ENZYMATIC REACTIONS IN NITROGEN FIXATION ARE  
 CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE  
 CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.  
 CC -1- CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP  
 CC = 8 oxidized ferredoxin + 2 NH(3) + 16 ADP + 16 phosphate.  
 CC -1- COFACTOR: BINDS ONE 4Fe-4S CLUSTER PER DIMER.  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE NIFH / BCLH / CHLL FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF003336; AAB61408.1; -;  
 CC HSP; P00459; IFF6.  
 CC HAMAP; MF\_00533; atypical; 1.  
 CC InterPro; IPR005977; NifH.  
 CC InterPro; IPR000392; NitrogenaseII.  
 CC Pfam; PF00142; fer4\_NifH; 1.  
 CC PRINTS; PRO0091; NITROGENASEII.  
 CC TIGRFAMS; TIGR01287; nifH; 1.  
 CC PROSITE; PS00746; NIFH\_FRXC\_1; 1.  
 CC PROSITE; PS00692; NIFH\_FRXC\_2; 1.  
 CC Oxidoreductase; Nitrogen fixation; Iron-sulfur; 4Fe-4S; ATP-binding.  
 FT NP BIND 45 52 ATP (POTENTIAL).  
 FT METAL 132 132 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 166 166 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT DOMAIN 1 37 PEST-LIKE; NOT FOUND IN OTHER NIFH.  
 SQ SEQUENCE 327 AA; D26946E425517D36 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 RELSRKI 94  
 DB 275 RELSRKI 281

RESULT 6  
 ID\_MC4R\_HUMAN STANDARD; PRT; 332 AA.  
 AC P32245; Q16317;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Melanocortin-4 receptor (MC4-R).  
 GN MC4R.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93315499; PubMed=8392067;  
 RA Gantz I., Miwa H., Konda Y., Shimoto Y., Tashiro T., Waston S.J.,  
 RA Delvalle J.;  
 RT "Molecular cloning, expression, and gene localization of a fourth  
 RT melanocortin receptor.";  
 RL J. Biol. Chem. 268:15174-15179 (1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95157557; PubMed=7854347;  
 RA Mountjoy K.G., Mortrud M.T., Low M.J., Simerly R.B., Cone R.D.;

RT "Localization of the melanocortin-4 receptor (MC4-R) in  
 RT neuroendocrine and autonomic control circuits in the brain."  
 RL Mol. Endocrinol. 8:1298-1308(1994).  
 RN [3]  
 RP VARIANTS OBESITY ARG-30; VAL-37; LEU-78; MET-112; TRP-165; SER-252 AND  
 RP THR-317, AND VARIANTS VAL-103 AND LEU-251.  
 RX MEDLINE=99213392; PubMed=10199800;  
 RA Hinney A., Schmidt A., Nottelmann K., Heibult O., Becker J., Ziegler A.,  
 RA Gerber G., Sina M., Gorg T., Mayer H., Stegried W., Fichter M.,  
 RA Renschmidt H., Hebebrand J.;  
 RT "Several mutations in the melanocortin-4 receptor gene including a  
 RT nonsense and a frameshift mutation associated with dominantly  
 RT inherited obesity in humans."  
 RL J. Clin. Endocrinol. Metab. 84:1483-1486(1999).  
 RN [4]  
 RP VARIANTS OBESITY SER-274.  
 RX MEDLINE=21336749; PubMed=11443223;  
 RA Mergen M., Mergen H., Ozata M., Onur R., Onur C.;  
 RA "A novel melanocortin 4 receptor (MC4R) gene mutation associated with  
 RA morbid obesity."  
 RL J. Clin. Endocrinol. Metab. 96:3448-3448(2001).  
 CC -1- FUNCTION: RECEPTOR SPECIFIC TO THE HEPTAPEPTIDE CORE COMMON TO  
 CC ADRENOCORTICOTROPIC HORMONE AND ALPHA-, BETA-, AND GAMMA-MSH.  
 CC THIS RECEPTOR IS MEDIATED BY G PROTEINS THAT STIMULATES ADENYLATE  
 CC CYCLASE.  
 CC SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: BRAIN, PLACENTAL, AND GUT TISSUES.  
 CC -1- DISEASE: Defects in MC4R are a cause of autosomal dominant  
 CC obesity [MIM:601665].  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L08603; AAA35791.1; --  
 DR EMBL; S72415; BAB33341.1; --  
 DR PIR; A57055; A57055.  
 DR Genew; HGNC:6932; MC4R.  
 DR MIN; 155541; --  
 DR MIN; 601665; --  
 DR GO; GO:0016021; C: integral to membrane; TAS.  
 DR GO; GO:0005886; C: plasma membrane; TAS.  
 DR GO; GO:0004977; F: melanocortin receptor activity; TAS.  
 DR GO; GO:0007631; P: feeding behavior; TAS.  
 DR GO; GO:0007388; P: G-protein signaling, coupled to cAMP nucleo. . . ; TAS.  
 DR GO; GO:0007582; P: physiological processes; TAS.  
 DR InterPro; IPR000276; GPCR\_Rhodopsin.  
 DR Pfam; PF00001; 7tm.1; 1.  
 DR PRINTS; PR00237; GPCRHHODOPSIN.  
 DR PROSITE; PS00237; G PROTEIN RECEPTOR FL 1; 1.  
 DR PROSITE; PS00262; G PROTEIN RECEPTOR FL 2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW phosphorylation; Lipoprotein; Palmitate; Polymorphism;  
 KW Disease mutation; Obesity.  
 FT DOMAIN 1 43 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 44 69 1 (POTENTIAL).  
 FT DOMAIN 70 81 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 82 106 2 (POTENTIAL).  
 FT DOMAIN 107 123 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 124 145 3 (POTENTIAL).  
 FT DOMAIN 146 165 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 166 186 4 (POTENTIAL).  
 FT DOMAIN 187 191 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 192 215 5 (POTENTIAL).  
 FT DOMAIN 216 248 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 249 271 6 (POTENTIAL).  
 FT DOMAIN 272 280 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 281 304 7 (POTENTIAL).

FT DOMAIN 305 332 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. . . ) (POTENTIAL).  
 FT CARBOHYD 17 17 N-LINKED (GLCNAC. . . ) (POTENTIAL).  
 FT CARBOHYD 26 26 N-LINKED (GLCNAC. . . ) (POTENTIAL).  
 FT LIPID 318 318 PALMITATE (POTENTIAL).  
 FT VARIANT 30 30 S -> R (IN OBESITY).  
 FT VARIANT 37 37 FTID=VAR\_010704.  
 FT VARIANT 37 37 D -> V (IN OBESITY).  
 FT VARIANT 78 78 P -> L (IN OBESITY).  
 FT VARIANT 103 103 FTID=VAR\_010705.  
 FT VARIANT 112 112 FTID=VAR\_010706.  
 FT VARIANT 165 165 FTID=VAR\_010707.  
 FT VARIANT 251 251 FTID=VAR\_010708.  
 FT VARIANT 252 252 FTID=VAR\_010709.  
 FT VARIANT 274 274 FTID=VAR\_010710.  
 FT VARIANT 317 317 FTID=VAR\_010711.  
 FT CONFLICT 169 169 FTID=VAR\_010712.  
 FT CONFLICT 332 332 FTID=VAR\_010713.  
 SQ SEQUENCE 332 AA; 36956 MW; 1C89752CF4DF30C CRC64;  
 Query Match 5.0%; Score 7; DB 1; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 52 VLPCTGA 58  
 Dd 228 VLPCTGA 234  
 RESULT 7  
 SYFA\_PSEAE STANDARD; PRT; 338 AA.  
 AC Q910A3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Phenylalanine-tRNA synthetase alpha chain (EC 6.1.1.20)  
 DE (Phenylalanine-tRNA ligase alpha chain) (PHERS).  
 GS PHER OR PA2740.  
 OS Pseudomonas aeruginosa.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 CC Pseudomonadaceae; Pseudomonas.  
 CX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Watrener P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.K., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Faulsen I.T.,  
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +  
 CC diphosphate + L-phenylalanyl-tRNA(Phe).  
 CC -1- COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).  
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
 CC Phe-tRNA synthetase alpha chain subfamily 1.  
 CC -----



\_\_\_\_\_

FT TURN 210 211  
FT STRAND 212 212  
FT STRAND 215 225  
FT TURN 226 227  
FT TURN 230 245  
FT TURN 247 248  
FT STRAND 251 255  
FT TURN 259 260  
FT STRAND 261 271  
FT TURN 272 275  
FT STRAND 276 286  
FT STRAND 288 300  
FT TURN 301 302  
FT STRAND 311 318  
FT TURN 319 327  
FT TURN 328 328  
FT TURN 332 337  
FT HELIX 340 343  
FT HELIX 344 346  
SQ SEQUENCE 350 AA; 39259 MW; DE02377FCD9A461 CRC64;  
  
Query Match 5.0%; Score 7; DB 1; Length 350;  
Best Local Similarity 100.0%; Pred.No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 36 ERLAMLR 42  
Db 320 ERLAMLR 326  
  
RESULT 9  
RL4A YEAST  
ID RL4A YEAST STANDARD; PRT; 361 AA.  
AC P10654;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE 60S ribosomal protein L2A (L2A) (RP2).  
GN RPL4A OR RPL2A OR RPL2 OR YBR031W OR YBR0315.  
OS Saccharomyces cerevisiae (Baker's Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88198160; PubMed=2834365;  
RA Presutti C., Lucicli A., Bozzoni I.;  
RT "Ribosomal protein L2 in Saccharomyces cerevisiae is homologous to  
RT ribosomal protein L1 in Xenopus laevis. Isolation and  
RT characterization of the genes."  
RL J. Biol. Chem. 263:6188-6192(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=S288c;  
RA Smits P.H.M., de Haan M., Maat C., Grivell L.A.;  
RT "The complete sequence of a 33 kb fragment on the right arm of  
RT chromosome II from Saccharomyces cerevisiae reveals 16 open reading  
RT frames, including ten new open reading frames, five previously  
RT identified genes and a homologue of the SCOL gene."  
RL Yeast 10:S75-360(1994).  
RN [3]  
RP L2 MRNA REGULATION.  
RX MEDLINE=91293097; PubMed=2065661;  
RA Presutti C., Ciafre S.-A., Bozzoni I.;  
RT "The ribosomal protein L2 in S. cerevisiae controls the level of  
RT accumulation of its own mRNA."  
RL EMBO J. 10:2215-2221(1991).  
RN [4]  
RP SEQUENCE OF 1-20.  
RX MEDLINE=92184799; PubMed=1544921;  
RA Takakura H., Tsunasawa S., Miyagi M., Warner J.R.;  
RT "NH2-terminal acetylation of ribosomal proteins of Saccharomycetes

RT cerevisiae.";  
RL J. Biol. Chem. 267:5442-5445(1992).  
CC -!- FUNCTION: PARTICIPATES IN THE REGULATION OF THE ACCUMULATION  
CC OF ITS OWN MRNA.  
CC -!- MISCELLANEOUS: THERE ARE TWO GENES FOR L4 IN YEAST.  
CC -!- SIMILARITY: BELONGS TO THE L4E FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X76078; CAA53687.1; -;  
DR EMBL; J03195; AAA34974.1; -;  
DR EMBL; Z35900; CAA84973.1; -;  
DR EIR; S45887; S45887.  
DR FDB; 1K5Y; 22-MAY-02.  
DR SGD; S000235; RPL4A.  
DR InterPro; IPR002136; Ribosomal L4/LiE.  
DR Pfam; PF00573; Ribosomal L4; 1.  
DR PROSITE; PS00939; RIBOSOMAL LiE; 1.  
KW Ribosomal protein; RNA-binding; Acetylation; Multigene family;  
KM 3D-structure.  
FT INIT\_MET 0 0  
FT MOD\_RES 1 1  
FT CONFLICT 37 37 V -> L (IN REF. 1).  
FT CONFLICT 143 143 K -> T (IN REF. 1).  
FT CONFLICT 156 156 E -> D (IN REF. 1).  
FT CONFLICT 223 223 G -> S (IN REF. 1).  
FT CONFLICT 240 240 G -> S (IN REF. 1).  
SQ SEQUENCE 361 AA; 38961 MW; 38272ACDADC8AA5F CRC64;  
  
Query Match 5.0%; Score 7; DB 1; Length 361;  
Best Local Similarity 100.0%; Pred.No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 58 ASAIAT 64  
Db 122 ASAIAT 128  
  
RESULT 10  
RL4B YEAST  
ID RL4B YEAST STANDARD; PRT; 361 AA.  
AC P49626;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE 60S ribosomal protein L4-B (L2B) (RP2).  
GN RPL4B OR RPL2B OR YBR012W OR YBR119.17.  
OS Saccharomyces cerevisiae (Baker's Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=S288c / AB972;  
RA Murphy L., Richards C., Gentles S., Harris D., Barrell B.G.,  
RA Rajandream M.A.;  
RN Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97051598; PubMed=8896275;  
RA Side L.G., Sander C., Prydz H.;  
RT "Sequencing and analysis of a 35.4 kb region on the right arm of  
RT chromosome IV from Saccharomycetes cerevisiae reveal 23 open reading  
RT frames."  
RL Yeast 12:1085-1090(1996).  
RN [3]  
RP SEQUENCE OF 1-199 FROM N.A.



RX MEDLINE=89096952; PubMed=3062369;  
RA Lucifoli A., Presutti C., Clafre S., Caffarelli E., Fragapane P.,  
RT Bozzoni I.;  
RA "Gene dosage alteration of L2 ribosomal protein genes in  
RT Saccharomyces cerevisiae: effects on ribosome synthesis.";  
RL Mol. Cell. Biol. 8:4792-4798(1988).  
RN [4].  
RP SEQUENCE OF 1-20.  
RX MEDLINE=92184799; PubMed=1544921;  
RA Takakura H., Tsunawasa S., Miyagi M., Warner J.R.;  
RA "NH2-terminal acetylation of ribosomal proteins of Saccharomyces  
RT cerevisiae.";  
RL J. Biol. Chem. 267:5442-5445(1992).  
RN [4].  
CC -!- FUNCTION: PARTICIPATES IN THE REGULATION OF THE ACCUMULATION  
CC OF ITS OWN MRNA.  
CC -!- MISCELLANEOUS: THERE ARE TWO GENES FOR L4 IN YEAST.  
CC -!- SIMILARITY: BELONGS TO THE L4E FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z48008; CAA88072.1; -;  
CC EMBL; X95966; CAA65204.1; -;  
CC EMBL; Z74308; CAA8832.1; -;  
CC EMBL; M22583; AAA34975.1; -;  
CC PIR; S50993; S50993.  
CC SGD; S0002419; RPL4B.  
CC InterPro; IPR002136; Ribosomal L4/L1e.  
CC Pfam; PF00573; Ribosomal L4; 1.  
CC PROSITE; PS00939; RIBOSOMAL L1e; 1.  
KW Ribosomal protein; RNA-binding; Acetylation; Multigene family.  
FT INIT MET 0 0  
FT MOD\_RES 1 1 ACETYLATION.  
FT CONFLICT 87 87 G -> R (IN REF. 3).  
SQ SEQUENCE 361 AA; 38931 MW; 38272ACD4DC8B62F CRC64;  
Query Match 5.0%; Score 7; DB 1; Length 361;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 58 ASATAAT 64  
DB 122 ASATAAT 128  
RESULT 11  
MK32 YEAST  
ID MK32 YEAST STANDARD; PRT; 363 AA.  
AC P23060;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE MAK32 protein.  
GN MAK32 OR YCR019W OR YCR19W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89131254; PubMed=3916862;  
RA Toh-E A., Sahashi Y.;  
RA "The PET18 locus of Saccharomyces cerevisiae: a complex locus  
RT containing multiple genes.";  
RL Yeast 1:159-171(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Feldmann H., Mannhaupt G., Vetter I.;

RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: NECESSARY FOR THE STRUCTURAL STABILITY OF L-A DOUBLE-  
CC STRANDED RNA-CONTAINING PARTICLES. NECESSARY FOR GROWTH AT 37-  
CC DEGREES CELSIUS AS WELL AS FOR MAINTENANCE OF THE KILLER PLASMID.  
CC -!- SIMILARITY: TO S-POMBE SPAC4G8.14C.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X59720; CAA42310.1; -;  
CC PIR; S19429; BVEYK2.  
CC SGD; S0000612; MAK32.  
CC GO; GO:0019048; P.virus-host interaction; IMP.  
CC InterPro; IPR002173; PfKB.  
CC Pfam; PF00294; pfKB; 1.  
CC CONFLICT 15 15 1 -> II (IN REF. 1).  
CC CONFLICT 82 82 MISSING (IN REF. 1).  
SQ SEQUENCE 363 AA; 40783 MW; CFB358F8ACF6EA4C CRC64;  
Query Match 5.0%; Score 7; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 29 LLRVKSS 35  
DB 158 LLRVKSS 164  
RESULT 12  
NODC\_RHILV  
ID NODC\_RHILV STANDARD; PRT; 424 AA.  
AC P04340;  
DT 20-MAR-1987 (Rel. 04, Created)  
DT 20-MAR-1987 (Rel. 04, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE N-acetylglucosaminyltransferase (EC 2.4.1.-) (Modulation protein C).  
GN NODC.  
OS Rhizobium leguminosarum (biovar viciae).  
OG Plasmid sym pRL1J1.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
OX NCBI\_TaxID=387;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=248;  
RX MEDLINE=85087952; PubMed=6514582;  
RA Rossen L., Johnston A.W.B., Downie J.A.;  
RT "DNA sequence of the Rhizobium leguminosarum nodulation genes nodAB  
RT and C required for root hair curling.";  
RL Nucleic Acids Res. 12:9497-9508(1984).  
RN [2]  
RP SEQUENCE OF 91-180 FROM N.A.  
RC STRAIN=USDA 2478;  
RX MEDLINE=95113787; PubMed=7814339;  
RA Ueda T., Suga Y., Yahiro N., Matsuguchi T.;  
RT "Phylogeny of Sym plasmids of rhizobia by PCR-based sequencing of a  
RT nodC segment.";  
RL J. Bacteriol. 177:468-472(1995).  
CC -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF NOD FACTOR, A SULFATED  
CC N-ACYL-BETA-1,4-TETRASACCHARIDE OF N-ACETYLGLUCOSAMINE WHICH  
CC INITIATES A SERIES OF EVENTS IN THE HOST PLANT SPECIES LEADING  
CC EVENTUALLY TO NODULATION.  
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).  
CC -!- SIMILARITY: BELONGS TO THE NODC/HAS FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)

CC -----  
 CC EMBL; Y00548; CA868619.1; -  
 CC EMBL; X01650; CA225814.1; -  
 CC EMBL; D28960; BAA06086.1; -  
 CC PIR; A03486; ZZZRCL.  
 CC InterPro; IPR004835; Fungi\_chitin\_syn.  
 CC InterPro; IPR001173; Glyco\_trans\_2.  
 CC Pfam; PF03142; Chitin synth 2; 1.  
 CC Pfam; PF00535; Glycosyltransf 2; 1.  
 CC Transferase; Glycosyltransferase; Nodulation; Membrane; Plasmid.  
 CC VARIANT 97 97 V -> E (IN STRAIN USDA 2478).  
 CC VARIANT 110 110 N -> K (IN STRAIN USDA 2478).  
 CC VARIANT 123 123 A -> A (IN STRAIN USDA 2478).  
 CC SEQUENCE 424 AA; 46255 MW; 1B1D2722A51A0FCF CRC64;

Query Match 5.0%; Score 7; Length 424;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATV 65  
 DB 259 SAIAATV 265

## RESULT 13

AM13 MYCLE  
 ID AM13 MYCLE STANDARD; PRT; 468 AA.  
 AC 033040;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative amidase amC (EC 3.5.1.4).  
 GN AMIC OR MLI596 OR MLCB250.65.  
 OS Mycobacterium leprae.  
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 CC NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=TN;  
 RC MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eiglmeter K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churche C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011(2001).

CC -1- CATALYTIC ACTIVITY: A monocarboxylic acid amide + H(2)O = a  
 CC monocarboxylate + NH(3).

CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)

CC EMBL; Z97369; CAB10659.1; -  
 CC EMBL; AL583922; CAC30547.1; -  
 CC PIR; F87108; F87108.  
 CC Leproma; MLI596; -  
 CC InterPro; IPR000120; Amidase.

DR Pfam; PF01425; Amidase; 1.  
 DR PROSITE; PS00571; AMIDASES; 1.  
 KW Hypothetical protein; Hydrolase; Complete proteome.  
 SQ SEQUENCE 468 AA; 51164 MW; A8E048D702F6A234 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 468;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 SPELREL 90  
 DB 269 SPELREL 275

## RESULT 14

PUR9 AQUAE  
 ID PUR9 AQUAE STANDARD; PRT; 506 AA.

AC 067775;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Bifunctional purine biosynthesis protein purH [includes:  
 DE Phosphoribosylamidoimidazolecarboxamide formyltransferase (EC 2.1.2.3)  
 DE (AICAR transferase); IMP cyclohydrolase (EC 3.5.4.10) (Inosinase)  
 DE (IMP synthetase) (ATIC)].  
 DE PURH OR AQ\_1983.  
 GN Aquifex aeolicus.

OS Aquifex aeolicus.  
 CC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 CC NCBI\_TaxID=63363;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VF5;  
 RC MEDLINE=98196666; PubMed=9537320;

RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aufay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus."

RL Nature 392:353-358(1998).

CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-  
 CC phospho-D-riboyl)imidazole-4-carboxamide = tetrahydrofolate + 5-  
 CC formamido-1-(5-phospho-D-riboyl)imidazole-4-carboxamide.

CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-phospho-D-  
 CC ribosyl)imidazole-4-carboxamide.

CC -1- PATHWAY: De novo purine biosynthesis; ninth step.  
 CC -1- PATHWAY: De novo purine biosynthesis; tenth step.

CC -1- DOMAIN: THE IMP CYCLOHYDROLASE ACTIVITY RESIDUES IN THE N-TERMINAL  
 CC REGION (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE PURH FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)

CC EMBL; A8E00765; AAC07734.1; -  
 CC PIR; C70468; C70468.

CC HAMAP; MF\_00139; -; 1.

CC InterPro; IPR002695; AICARFT IMPChas.

CC InterPro; IPR004362; MGS like.

CC Pfam; PF01808; AICARFT IMPChas; 1.

CC Pfam; PF02142; MGS; 1.

CC ProDom; PD004666; AICARFT IMPChas; 1.

CC TIGRFAMs; TIGR00355; purH; 1.

CC Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme;

CC Complete proteome.

CC SEQUENCE 506 AA; 56677 MW; 46998C8B1B1F694 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 506;  
 Best Local Similarity 100.0%; Pred. No. 24;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 LRELSRK 93  
Db 335 LRELSRK 341

RESULT 15  
TXB2\_MOUSE  
ID TXB2\_MOUSE STANDARD; PRT; 701 AA.  
AC Q60707;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE T-box transcription factor TBX2 (T-box protein 2).  
GN TBX2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND DEVELOPMENTAL EXPRESSION.  
RC TISSUE=Embryo;  
RX MEDLINE=95004605; PubMed=7920656;  
RA Bollag R.J., Siegfried Z., Cebra-Thomas J.A., Garvey N., Davison E.M.,  
Silver L.M.;  
RT "An ancient family of embryonically expressed mouse genes sharing a  
conserved protein motif with the T locus.";  
RL Nat. Genet. 7:383-389(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20357305; PubMed=10770922;  
RA Carreira S., Liu B., Goding C.R.;  
RT "The gene encoding the T-box factor Tbx2 is a target for the  
microphthalmia-associated transcription factor in melanocytes.";  
RL J. Biol. Chem. 275:21920-21927(2000).  
RN [3]  
RP DEVELOPMENTAL EXPRESSION  
RX MEDLINE=97006694; PubMed=8953987;  
RA Chapman D.L., Garvey N., Hancock S., Alexiou M., Agulnik S.I.,  
Gibson-Brown J.J., Cebra-Thomas J., Bollag R.J., Silver L.M.,  
Papalioannou V.E.;  
RT "Expression of the T-box family genes, Tbx1-Tbx5, during early mouse  
development.";  
RL Dev. Dyn. 206:379-390(1996).  
CC -!- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES  
CC REQUIRED FOR MESODERM DIFFERENTIATION. PROBABLY PLAYS A ROLE IN  
CC LIMB PATTERN FORMATION.  
CC -!- TISSUE SPECIFICITY: IN ADULTS, HIGHEST LEVELS IN LUNG. ALSO FOUND  
CC IN HEART, KIDNEY, AND OVARY.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSION FIRST OBSERVED AT DAY 9.5 IN  
CC THE OTIC AND OPTIC VESICLES AND IN THE FACIAL REGION AT DAY  
CC 12.5, EXPRESSED IN THE TRIGEMINAL GANGLIA, FACIAL REGIONS, RETINA  
CC AND LIMB BUD MESENCHYME. IN LATER STAGES, FOUND IN EAR PINNAE,  
CC THE MILK LINE, LUNG MESENCHYME, BODY WALL, GENITAL RIDGE AND  
CC DEVELOPING NERVOUS SYSTEM.  
CC -!- SIMILARITY: Contains 1 T-box domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U15566; AAC52697.1; -.  
CC EMBL; AF244917; AAF90050.1; -.  
CC PIR; S46458; S46458.  
CC HSSP; P24781; 1XBR.  
CC TRANSFAC; T04358; -.  
CC MGD; MGI:98494; Tbx2.

GO; GO:0005667; C:transcription factor complex; IDA.  
GO; GO:0005515; F:protein binding activity; IPI.  
GO; GO:0016564; P:transcriptional repressor activity; IDA.  
GO; GO:0000122; P:negative regulation of transcription from P...; IDA.  
InterPro; IPR001699; TF\_T-box.  
Pfam; PF00907; T-box; 1.  
PRINTS; PR00937; TBOX..  
SMART; SMO0425; TBOX; 1.  
PROSITE; PS01283; TBOX\_1; 1.  
PROSITE; PS01264; TBOX\_2; 1.  
PROSITE; PS0252; TBOX\_3; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein;  
Developmental protein.  
FT DOMAIN 48 63 POLY-ALA.  
FT DNA BIND 104 277 T-BOX.  
FT DOMAIN 572 580 POLY-ALA.  
FT DOMAIN 586 594 POLY-ALA.  
SQ SEQUENCE 701 AA; 74244 MW; 8D90ED6DA32B3859 CRC64;  
Query Match 5.0%; Score 7; DB 1; Length 701;  
Best Local Similarity 100.0%; Pred No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 HLRLSKS 134  
Db 78 HLRLSKS 84

RESULT 16  
TXB2\_HUMAN  
ID TXB2\_HUMAN STANDARD; PRT; 702 AA.  
AC Q13207; Q16424;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE T-box transcription factor TBX2 (T-box protein 2).  
GN TBX2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Fetal kidney;  
RX MEDLINE=96015055; PubMed=8530034;  
RA Campbell C., Goodrich K., Casey G., Beatty B.;  
RT "Cloning and mapping of a human gene (TBX2) sharing a highly conserved  
protein motif with the Drosophila omb gene.";  
RL Genomics 28:255-260(1995).  
RN [2]  
RP SEQUENCE OF 152-245 FROM N.A.  
RX TISSUE=Fetal kidney;  
RX MEDLINE=96169568; PubMed=8597636;  
RA Law D.J., Gebuhr T., Garvey N., Agulnik S.I., Silver L.M.;  
RT "Identification, characterization, and localization to chromosome  
17q21-22 of the human TBX2 homolog, member of a conserved  
developmental gene family.";  
RL Mamm Genome 6:793-797(1995).  
CC -!- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES  
CC REQUIRED FOR MESODERM DIFFERENTIATION. PROBABLY PLAYS A ROLE IN  
CC LIMB PATTERN FORMATION.  
CC -!- SUBCELLULAR LOCATION: Nuclear (potential).  
CC -!- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN ADULT IN KIDNEY, LUNG,  
CC AND PLACENTA. WEAK EXPRESSION IN HEART AND OVARY.  
CC -!- SIMILARITY: Contains 1 T-box domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC EMBL; U28049; AAB73861.1; -  
DR EMBL; S81264; AAB36216.1; -  
DR PIR; G01840; G01840.  
DR HSSP; P24781; IXB8.  
DR TRANSFAC; T04351; -  
DR Genew; HGNC:11597; TBX2.  
DR MIM; 600747; -  
DR InterPro; IPR001699; TF\_I-box.  
DR Pfam; PF00907; T-box\_1.  
DR PRINTS; PR00937; TBOX.  
DR SMART; SM00445; TBOX; 1.  
DR PROSITE; PS01283; TBOX\_1; 1.  
DR PROSITE; PS01264; TBOX\_2; 1.  
DR PROSITE; PS0252; TBOX\_3; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein;  
Developmental protein.  
FT DOMAIN 48 63 POLY-ALA.  
FT DNA\_BIND 104 277 T-BOX.  
FT DOMAIN 507 517 POLY-GLY.  
FT DOMAIN 571 579 POLY-ALA.  
FT DOMAIN 585 593 POLY-ALA.  
FT CONFLICT 155 155 Y -> D (IN REF. 2).  
FT CONFLICT 165 168 AGKA -> TDKT (IN REF. 2).  
SQ SEQUENCE 702 AA; 74194 MW; C6477134C69D7C2C CRC64;  
Query Match 5.0%; Score 7; DB 1; Length 702;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 128 HLRLSKS 134  
Db 78 HLRLSKS 84  
RESULT 17  
PTIP\_ECOLI STANDARD; PRT; 748 AA.  
AC P37177;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Phosphoenolpyruvate-protein phosphotransferase ptsp (EC 2.7.3.9)  
DE (Phosphotransferase system, enzyme I) (Enzyme I-Ntr).  
GN PTSP OR P2829.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RA MEDLINE=97426617; PubMed=9278503;  
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE OF 656-748 FROM N.A.  
RX MEDLINE=95204358; PubMed=7896715;  
RA Gan K., Sankaran K., Williams M.G., Aldea M., Rudd K.E., Kushner S.R.,  
Wu H.C.;  
RT "The umpa gene of Escherichia coli encodes  
phosphatidylglycerol:phosphoprotein diacylglyceryl transferase (1gt)  
and regulates thymidylate synthase levels through translational  
coupling.";  
RL J. Bacteriol. 177:1879-1882(1995).  
RN [3]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=97128775; PubMed=8973315;  
RA Reizer J., Reizer A., Merrick M.J., Plunkett G. III, Rose D.J.,

RA Saier M.H. Jr.;  
RT "Novel phosphotransferase-encoding genes revealed by analysis of the  
Escherichia coli genome: a chimeric gene encoding an Enzyme I  
homologue that possesses a putative sensory transduction domain.";  
RL Gene 181:103-108(1996).  
CC -I- FUNCTION: COULD FUNCTION IN TRANSCRIPTIONAL REGULATION OF SIGMA-54  
DEPENDENT OPERONS IN CONJUNCTION WITH THE NPR (PTSO) AND IIA-NTR  
(PTSN) PROTEINS THEREBY PROVIDING A LINK BETWEEN CARBON AND  
NITROGEN ASSIMILATORY PATHWAYS.  
CC -I- CATALYTIC ACTIVITY: Phosphoenolpyruvate + protein L-histidine =  
pyruvate + protein N(pi)-phospho-L-histidine.  
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -I- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.  
CC -I- SIMILARITY: IN THE N-TERMINAL, TO NIFA PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; U29581; AAB40476.1; -  
DR EMBL; AE000366; AAC75868.1; -  
DR PIR; U12289; AAA69023.1; -  
DR EcoGene; EG12188; dtsp.  
DR InterPro; IPR003018; GAP.  
DR InterPro; IPR006318; PEP\_P\_trans.  
DR InterPro; IPR000121; PEP\_utilizers.  
DR Pfam; PF01590; GAP; 1.  
DR Pfam; PF00391; PEP-utilizers; 1.  
DR Pfam; PF02896; PEP-utilizers; 1.  
DR ProDom; PD000940; PEP\_utilizers; 1.  
DR SMART; SM00065; GAP; 1.  
DR TIGRfam; TIGR01417; PTS\_I\_fam; 1.  
DR PROSITE; PS00370; PEP\_ENZYMES\_PHOS\_SITE; 1.  
DR PROSITE; PS00742; PEP\_ENZYMES\_2; 1.  
KW Phosphotransferase system; Transferase; Kinase; Sugar transport;  
KW Phosphorylation; Complete proteome.  
FT DOMAIN 1 127 NIFA-LIKE.  
FT DOMAIN 128 170 LINKER.  
FT DOMAIN 171 748 ENZYME 1 DOMAIN.  
FT MOD\_RES 356 356 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 748 AA; 83715 MW; AC7137BD0AEBBF01 CRC64;  
Query Match 5.0%; Score 7; DB 1; Length 748;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 39 AMLRALA 45  
Db 646 AMLRALA 652  
RESULT 18  
ELS\_MOUSE STANDARD; PRT; 860 AA.  
AC P54320;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Elastin precursor (Tropoelastin).  
GN ELN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BA1B/C; TISSUE=Lung;  
RA MEDLINE=95130069; PubMed=7829060;

RA Wydner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;  
 RT "Use of an intron polymorphism to localize the tropoelastin gene to  
 RT mouse chromosome 5 in a region of linkage conservation with human  
 RT chromosome 7";  
 RL Genomics 23:125-131(1994).  
 CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND  
 CC NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.  
 CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER  
 CC INTO AN EXTENSIBLE 3D NETWORK.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=8;  
 CC Comment=Experimental confirmation may be lacking for some  
 CC isoforms;  
 CC Name=1;  
 CC IsoId=Q99372-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q99372-2; Sequence=VSP\_004244;  
 CC Name=3;  
 CC IsoId=Q99372-3; Sequence=VSP\_004245;  
 CC Name=4;  
 CC IsoId=Q99372-4; Sequence=VSP\_004246;  
 CC Name=5;  
 CC IsoId=Q99372-5; Sequence=VSP\_004244, VSP\_004245;  
 CC Name=6;  
 CC IsoId=Q99372-6; Sequence=VSP\_004245, VSP\_004246;  
 CC Name=7;  
 CC IsoId=Q99372-7; Sequence=VSP\_004244, VSP\_004246;  
 CC Name=8;  
 CC IsoId=Q99372-8; Sequence=VSP\_004244, VSP\_004245, VSP\_004246;  
 CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U08210; AAA80155.1; -;  
 CC PIR; A55721; EAMS.  
 CC GPD; MGI:95317; Eln.  
 CC InterPro; IPR003979; tropoelastin.  
 CC PRINTS; PR01500; TROPOELASTIN.  
 CC Structural protein; Repeat; Signal; Connective tissue.  
 CC SIGNAL 1 27 POTENTIAL.  
 CC CHAIN 28 860 ELASTIN.  
 CC DISULFID 850 855 BY SIMILARITY.  
 CC SEQUENCE 860 AA; 71955 MW; 0C0BESA1EDD7F1 CRC64;  
 CC  
 CC Query Match 5.0%; Score 7; DB 1; Length 860;  
 CC Best Local Similarity 100.0%; Pred. No. 39;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 52 VLPGTGA 58  
 CC DB 215 VLPGTGA 221  
 CC  
 CC RESULT 19  
 CC ELS\_RAT  
 CC ID\_ELS\_RAT STANDARD; PRT; 864 AA.  
 CC AC Q99372;  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 CC DE Elastin precursor (tropoelastin) (Fragment).  
 CC GN ELN.  
 CC OS Rattus norvegicus (Rat).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC OX NCBI\_TaxID=10116;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 CC RA MEDLINE=91104868; PubMed=1702999;  
 CC RA Pierce R.A., Deak S.B., Stolle C.A., Boyd C.D.;  
 CC RT "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";  
 CC RL Biochemistry 29:9677-9683(1990).  
 CC RN [2]  
 CC RP SEQUENCE OF 781-864 FROM N.A.  
 CC RA MEDLINE=88330868; PubMed=2971041;  
 CC RA Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;  
 CC RT "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";  
 CC RL J. Biol. Chem. 263:13504-13507(1988).  
 CC RN [3]  
 CC RP SEQUENCE OF 264-533 AND 558-864 FROM N.A., AND ALTERNATIVE SPLICING.  
 CC RA MEDLINE=92241859; PubMed=1572637;  
 CC RA Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;  
 CC RT "Elements of the rat tropoelastin gene associated with alternative  
 CC RT splicing";  
 CC RL Genomics 12:651-658(1992).  
 CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND

CC NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.  
 CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER  
 CC INTO AN EXTENSIBLE 3D NETWORK.  
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=8;  
 CC Comment=Experimental confirmation may be lacking for some  
 CC isoforms;  
 CC Name=1;  
 CC IsoId=Q99372-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q99372-2; Sequence=VSP\_004244;  
 CC Name=3;  
 CC IsoId=Q99372-3; Sequence=VSP\_004245;  
 CC Name=4;  
 CC IsoId=Q99372-4; Sequence=VSP\_004246;  
 CC Name=5;  
 CC IsoId=Q99372-5; Sequence=VSP\_004244, VSP\_004245;  
 CC Name=6;  
 CC IsoId=Q99372-6; Sequence=VSP\_004245, VSP\_004246;  
 CC Name=7;  
 CC IsoId=Q99372-7; Sequence=VSP\_004244, VSP\_004246;  
 CC Name=8;  
 CC IsoId=Q99372-8; Sequence=VSP\_004244, VSP\_004245, VSP\_004246;  
 CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; M60647; AAA42269.1; -;  
 CC EMBL; J04035; AAA42268.1; -;  
 CC EMBL; M86372; AAA42271.1; -;  
 CC EMBL; M86355; AAA42271.1; JOINED.  
 CC EMBL; M86363; AAA42271.1; JOINED.  
 CC EMBL; M86364; AAA42271.1; JOINED.  
 CC EMBL; M86366; AAA42271.1; JOINED.  
 CC EMBL; M86371; AAA42271.1; JOINED.  
 CC EMBL; M86376; AAA42272.1; -;  
 CC EMBL; M86373; AAA42272.1; JOINED.  
 CC EMBL; M86375; AAA42272.1; JOINED.  
 CC PIR; A36106; EART.  
 CC InterPro; IPR003979; tropoelastin.  
 CC PRINTS; PR01500; TROPOELASTIN.  
 CC Structural protein; Connective tissue; Repeat; Signal;  
 CC Alternative splicing.  
 CC NON\_TER 1 1  
 CC SIGNAL <1 21 BY SIMILARITY.  
 CC CHAIN 22 864 ELASTIN.  
 CC DISULFID 854 859 BY SIMILARITY.  
 CC VARSPLIC 263 307 Missing (in isoform 2, isoform 5, isoform  
 CC 7 and isoform 8).  
 CC FT FTID=VSP\_004244.  
 CC FT FTID=VSP\_004245.  
 CC FT FTID=VSP\_004246.  
 CC FT FTID=VSP\_004245.  
 CC FT FTID=VSP\_004246.  
 CC FT FTID=VSP\_004246.  
 CC SEQUENCE 864 AA; 72786 MW; 456894BB09E79FD4 CRC64;  
 CC  
 CC Query Match 5.0%; Score 7; DB 1; Length 864;  
 CC Best Local Similarity 100.0%; Pred. No. 39;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 52 VLPGTGA 58  
 CC DB 199 VLPGTGA 205

RESULT 20  
GCSP\_PEA  
ID GCSP\_PEA STANDARD; PRT; 1057 AA.  
AC P26969;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Glycine dehydrogenase [decarboxylating], mitochondrial precursor  
DE (EC 1.4.4.2) (Glycine decarboxylase) (Glycine cleavage system P-protein)  
DE GCSP OR GDCP.  
GN Pisum sativum (Garden pea).  
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.  
OX NCBI\_TaxID=3888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Birte; TISSUE=leaf;  
RX MEDLINE=92184787; PubMed=1347530;  
RA Turner S.R., Ireland R., Rawsthorne S.;  
RT "Cloning and characterization of the P subunit of glycine  
RT decarboxylase from pea (Pisum sativum).";  
RL J. Biol. Chem. 267:5355-5360(1992).  
RN [2]  
RP SEQUENCE OF 905-1057 FROM N.A.  
RC STRAIN=cv. Alaska;  
RA Shah K.S., Kim Y., Oliver D.J.;  
RL Submitted (AUG-1990) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: The glycine cleavage system catalyzes the degradation of  
CC glycine. The P protein binds the alpha-amino group of glycine  
CC through its pyridoxal phosphate cofactor; CO(2) is released and  
CC the remaining methylene moiety is then transferred to the  
CC liposamide cofactor of the H protein.  
CC -!- CATALYTIC ACTIVITY: Glycine + lipoylprotein = S-  
CC aminomethylidihydrolipoylprotein + CO(2).  
CC -!- COFACTOR: Pyridoxal phosphate.  
CC -!- SUBUNIT: HOMODIMER. THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF  
CC FOUR PROTEINS: P, T, L, AND H.  
CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
CC -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X59773; CAA42443.1; -;  
DR EMBL; X54377; CAA38252.1; -;  
DR PIR; A42109; A42109.  
DR InterPro: IPR003437; GDC-P.  
DR Pfam: PF02347; GDC-P; 1.  
DR TIGRfam: TIGR00461; gcvp; 1.  
KW Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide.  
FT TRANSIT 1 86 MITOCHONDRION (POTENTIAL).  
FT CHAIN 87 1057 GLYCINE DEHYDROGENASE (DECARBOXYLATING).  
FT BINDING 792 792 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
FT CONFLICT 905 905 I -> Y (IN REF. 2).  
FT CONFLICT 919 919 P -> A (IN REF. 2).  
SQ SEQUENCE 1057 AA; 114686 MW; 2FEA58E9A2AC47 CRC64;  
Query Match 5.0%; Score 7; DB 1; Length 1057;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 ESTSTTT 8  
DB 27 ESTSTTT 33

RESULT 21  
RAD9\_YEAST  
ID RAD9\_YEAST STANDARD; PRT; 1309 AA.  
AC P14737; Q04920;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE DNA repair protein RAD9.  
DE RAD9 OR YDR217C OR YD9934.02C.  
GN Saccharomyces cerevisiae (Baker's yeast).  
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89313732; PubMed=2664461;  
RA Schiestl R.H., Reynolds P., Prakash S., Prakash L.;  
RT "Cloning and sequence analysis of the Saccharomyces cerevisiae RAD9  
RT gene and further evidence that its product is required for cell cycle  
RT arrest induced by DNA damage.";  
RL Mol. Cell. Biol. 9:1882-1896(1989).  
RN [2]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=91061763; PubMed=2247073;  
RA Weinert T.A., Hartwell L.H.;  
RT "Characterization of RAD9 of Saccharomyces cerevisiae and evidence  
RT that its function acts posttranslationally in cell cycle arrest after  
RT DNA damage.";  
RL Mol. Cell. Biol. 10:6554-6564(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RX PHOSPHORYLATION, AND INTERACTION WITH RAD53.  
RX MEDLINE=98429491; PubMed=9755168;  
RA Vieland J.E., Gilbert C.S., Green C.M., Lowndes N.F.;  
RT "The budding yeast Rad9 checkpoint protein is subjected to  
RT Mec1/Tel1-dependent hyperphosphorylation and interacts with Rad53  
RT after DNA damage.";  
RL EMBO J. 17:5679-5688(1998).  
CC -!- FUNCTION: ESSENTIAL FOR CELL CYCLE ARREST AT THE G2 STAGE  
CC FOLLOWING DNA DAMAGE BY X-IRRADIATION OR INACTIVATION OF DNA  
CC LIGASE.  
CC -!- SUBUNIT: Physically associates with RAD53.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: Contains 1 BRCT domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; M26049; AAA34954.1; -;  
DR EMBL; Z48612; CAA88497.1; -;  
DR PIR; S59424; BVBYD9.  
DR PDB; 1FHR; 18-OCT-00.  
DR PDB; 1J4K; 05-DEC-01.  
DR PDB; 1J4L; 05-DEC-01.  
DR PDB; 1J4P; 05-DEC-01.  
DR PDB; 1J4Q; 05-DEC-01.  
DR PDB; 1K2M; 05-DEC-01.  
DR PDB; 1K2N; 05-DEC-01.  
DR PDB; 1K3N; 05-DEC-01.  
DR PDB; 1K3Q; 05-DEC-01.  
DR SGD; S0002625; RAD9.  
DR GO; GO:0005634; C:nucleus; IC.  
DR GO; GO:0005515; P:protein binding activity; IDA.  
DR GO; GO:0000077; P:DNA damage checkpoint; IMP.

DR GO; GO:0006289; P:nucleotide-excision repair; IMP.  
 DR GO; GO:0000074; P:regulation of cell cycle; IGI.  
 DR InterPro; IPR001357; BRCT.  
 DR Pfam; PF00533; BRCT; 1.  
 DR SMART; SM00292; BRCT; 1.  
 DR PROSITE; PS00172; BRCT; 1.  
 KW Cell cycle; DNA damage; DNA replication inhibitor; Nuclear protein;  
 KW Phosphorylation; 3D-structure.  
 FT DOMAIN 994 1122  
 FT CONFLICT 433 433 C -> S (IN REF. 3).  
 SQ SEQUENCE 1309 AA; 148413 MW; 6B77D39A95021F84 CRC64;

Query Match 5.0%; Score 7; DB 1; Length 1309;  
 Best Local Similarity 100.0%; Pred.No.58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 127 RHLRSLK 133  
 DB 1082 RHLRSLK 1088

RESULT 22  
 RPOP\_HALN1 STANDARD; PRT; 44 AA.  
 AC Q9HSG9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA-directed RNA polymerase subunit P [EC 2.7.7.6].  
 GN RPOP OR VNG0237H.  
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OX NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2050483; PubMed=11016950;  
 RA Ng M.V., Kennedy S.P., Maniatis G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Shrogha J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,  
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Beck R.F., Pohlshocher M., Spudis J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RA "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
 of DNA into RNA using the four ribonucleoside triphosphates as  
 substrates.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 {RNA}(N).  
 CC -!- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPOP / EUKARYOTIC RPOP10  
 RNA POLYMERASE SUBUNIT FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; AE004988; AAG18837.1; -.  
 DR PIR; A84184; A84184.  
 DR HAMAP; MF 00615; -; 1.  
 DR InterPro; IPR006591; RNA\_pol\_Rbp10.  
 DR SMART; SM00659; RPOPICK; 1.  
 KW Transferase; DNA-directed RNA polymerase; Transcription;  
 KW Metal-binding; Zinc-finger; Complete proteome.  
 FT ZN FING 5 25 C4-TYPE (POTENTIAL).  
 SQ SEQUENCE 44 AA; 5179 MW; C175E83AA4CADFAC CRC64;

Query Match 4.3%; Score 6; DB 1; Length 44;  
 Best Local Similarity 100.0%; Pred.No.28;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 48 CGHRVL 53  
 DB 25 CGHRVL 30

RESULT 23  
 ATP8\_ASTPE STANDARD; PRT; 54 AA.  
 AC Q3822;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DE 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L).  
 GN MTATP8 OR ATP8  
 OS Asterina pectinifera (Starfish).  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;  
 OC Asteroidea; Valvatacea; Valvatida; Asterinidae; Asterina.  
 OX NCBI\_TaxID=7594;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RX MEDLINE=95402698; PubMed=7672576;  
 RA Asakawa S., Himeno H., Miura K.-I., Watanabe K.;  
 RA "Nucleotide sequence and gene organization of the starfish Asterina  
 pectinifera mitochondrial genome";  
 RL Genetics 140:1047-1060(1995).  
 CC -!- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT  
 (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.  
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +  
 H(+)(Out).  
 CC -!- SUBCELLULAR LOCATION: Membrane-bound.  
 CC -!- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; D16387; BA03883.1; -.  
 DR PIR; S70600; S70600.  
 DR InterPro; IPR001421; ATPase8\_mit.  
 DR Pfam; PF00895; ATP-synt\_8; 1.  
 KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.  
 FT TRANSMEM 8 28 POTENTIAL.  
 SQ SEQUENCE 54 AA; 6241 MW; 9EABDACEB9CDF5F1 CRC64;

Query Match 4.3%; Score 6; DB 1; Length 54;  
 Best Local Similarity 100.0%; Pred.No.34;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 TSTTTN 9  
 DB 39 TSTTTN 44

RESULT 24  
 RKL4\_OENAM STANDARD; PRT; 74 AA.  
 ID RKL4\_OENAM  
 AC P42340;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Chloroplast 50S ribosomal protein L14 (Fragments).  
 GN RPL14.  
 OS Oenothera ammodendron (Evening primrose).



```

DR SMART; SW00389; H0X; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Nuclear protein; Multigene family.
FT NON_TER 1 1 ELK DOMAIN.
FT FT 2 25 HOMEBOX (TALE-TYPE).
FT DNA_BIND 26 88 POLY-LYS.
FT FT 27 32
FT DOMAIN 27 32
FT NON_TER 88 88
FT FT 88 88
SQ SEQUENCE 88 AA; 10789 MW; CEBFEDD754A024C2 CRC64;

Query Match 4.3%; Score 6; DB 1; Length 88;
Best Local Similarity 100.0%; Pred.No. 54;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 RELSRK 93
DB 22 RELSRK 27

RESULT 26
MMOD METTR STANDARD; PRT; 102 AA.
AC Q5352;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Methane monooxygenase component D.
GN MMOD.
OS Methylosinus trichosporium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Methylcytochrome; Methylosinus.
OX NCBI_TaxID=426;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OB3b;
RX MEDLINE=92153031; PubMed=1785954;
RA Cardy D.L.N., Laidler V., Salmond G.P.C., Murrell J.C.;
RT "The methane monooxygenase gene cluster of Methylosinus trichosporium:
RT cloning and sequencing of the mmoC gene.";
RL Arch. Microbiol. 156:477-483(1991).
CC -1- FUNCTION: Not known.
CC -1- SUBUNIT: The soluble methane monooxygenase (sMMO) consists of four
CC components A/MMOH (composed of alpha/mmoX, beta/mmoY and
CC gamma/mmoZ), B/MMOB (mmoB), C/MMOR (mmoC) and D/MMOD (mmoD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S81887; AAB21392.1; --
CC PIR; B48360; B48360.
CC Oxidoreductase; Monooxygenase.
SQ SEQUENCE 102 AA; 11926 MW; 055307622A09409D CRC64;

Query Match 4.3%; Score 6; DB 1; Length 102;
Best Local Similarity 100.0%; Pred.No. 62;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 SQESAR 107
DB 49 SQESAR 54

RESULT 27
ACCD ORISA
ID ACCD ORISA STANDARD; PRT; 106 AA.
AC P12218;
DT 01-OCT-1989 (Rel. 12, Created)

```



DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
DE (EC 6.4.1.2) (ACCASE beta chain).  
GN ACCD OR YCF11.  
OS Oryza sativa (Rice).  
OS Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OC NCBI\_TaxID=4530;  
RN NCBI\_SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RX MEDLINE=89364698; PubMed=2770692;  
RA Hiratauka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M.,  
RA Mori M., Kondo C., Honji Y., Sun C.-R., Meng B.-Y., Li Y.-Q.,  
RA Kanno A., Nishikawa Y., Hirai A., Shinozaki K., Sugita M.,  
RT "The complete sequence of the rice (*Oryza sativa*) chloroplast genome:  
RT intermolecular recombination between distinct tRNA genes accounts for  
RT a major plastid DNA inversion during the evolution of the cereals.";  
RL Mol. Gen. Genet. 217:185-194(1989).  
CC -|- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) = ADP + phosphate  
CC + malonyl-CoA.  
CC -|- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.  
CC -|- CAUTION: CORRESPONDS TO THE C-TERMINAL PART OF OTHER PLANTS  
CC ACCD.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X15901; CAA33956.1; -.  
DR PIR; J00234; J00234.  
DR Granene; P12218; -.  
KW Fatty acid biosynthesis; Ligase; Chloroplast.  
SQ SEQUENCE 106 AA; 12455 MW; D0410243163BFF2E CRC64;  
-----  
Query Match 4.3%; Score 6; DB 1; Length 106;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 80 QSTKSP 85  
Db 64 QSTKSP 69  
-----  
RESULT 28  
SSI\_STRGI STANDARD; PRT; 107 AA.  
AC P28592;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE Alkaline protease inhibitor 2C' (API-2C').  
OS Streptomyces griseoflavus.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomycetes.  
OC NCBI\_TaxID=29305;  
RN NCBI\_SEQUENCE FROM N.A.  
RC STRAIN=K. Ueyeda M., Shibata M.;  
RX MEDLINE=45629634(1981).  
RA (b and c).";  
RT Agric. Biol. Chem. 45:629-634(1981).  
CC -|- FUNCTION: STRONG INHIBITORY ACTIVITY TOWARD SUBTILISIN.  
CC -|- SUBUNIT: Homodimer.  
CC -|- SUBCELLULAR LOCATION: Secreted.  
CC -|- SIMILARITY: BELONGS TO THE SERINE PROTEASE INHIBITORS SSI FAMILY.

DR PIR; P01261; P01261.  
DR HSP; P01006; 2SIC.  
DR InterPro; IPR000691; Strep\_subt\_inhib.  
DR Pfam; PF00720; SSI; 1.  
DR PRINTS; PR00234; SSBTLINHETR.  
DR ProDom; PD004028; Strep\_subt\_inhib; 1.  
DR PROSITE; PS00999; SSI; 1.  
KW Serine protease inhibitor.  
FT DISULFID 29 44 BY SIMILARITY.  
FT DISULFID 65 95 BY SIMILARITY.  
FT ACT SITE 67 68 REACTIVE BOND.  
SQ SEQUENCE 107 AA; 10972 MW; 76043BA9F876D1CA CRC64;  
-----  
Query Match 4.3%; Score 6; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 62 AATVTP 67  
Db 16 AATVTP 21  
-----  
RESULT 29  
YHEA\_RHOCA STANDARD; PRT; 124 AA.  
AC P29952;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DE Hypothetical 12.7 kDa protein in hella 5' region (ORF124).  
OS Rhodospirillum rubrum (Rhodospirillum rubrum).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
OC Rhodobacteraceae; Rhodobacter.  
OC NCBI\_TaxID=1061;  
RN NCBI\_SEQUENCE FROM N.A.  
RC STRAIN=SB1003 / St Louis;  
RX MEDLINE=92146961; PubMed=1310666;  
RA Beckman D.L., Trawick D.R., Kranz R.G.;  
RA "Bacterial cytochromes c biogenesis.";  
RL Genes Dev. 6:268-283(1992).  
CC -|- FUNCTION: NOT REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X63462; CAA45060.1; -.  
DR PIR; S23662; S23662.  
DR InterPro; IPR006729; DUF598.  
DR Pfam; PF04635; DUF598; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 124 AA; 12711 MW; 1CA4FDFD6F77DFCE CRC64;  
-----  
Query Match 4.3%; Score 6; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 41 LRLAAG 46  
Db 52 LRLAAG 57  
-----  
RESULT 30  
ABAL\_TRIAB STANDARD; PRT; 131 AA.  
ID ABAL\_TRIAB  
AC P8111;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Alboaggregin A subunit 1.  
OS Trimeresurus albolabris (White-lipped pit viper).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosaurs; Squamata; Scleroglossa; Serpentes; Colubroidea;  
OC Viperidae; Crotalinae; Trimeresurus.  
OX NCBI\_TaxID=8765;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Venom;  
RX MEDLINE=98189535; PubMed=9531050;  
RA Kowalek M.A., Tan L., Holt J.C., Peng M., Karczewski J.,  
RA Calvete J.J., Nowiarski S.,  
RT "Alboaggregins A and B. Structure and interaction with human  
RT platelets."  
RL Thromb. Haemost. 79:609-613(1998).  
CC -1- FUNCTION: Binds to platelet GPIb/IX receptor system and stimulates  
CC agglutination.  
CC -1- SUBUNIT: Heterotrimer of the subunits 1, 2, 3 and 4,  
CC disulfide-linked.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.  
DR HSSP; P23806; 11XX.  
DR InterPro; IPR001304; Lectin\_C.  
DR Pfam; PF00059; lectin\_C; 1.  
DR SMART; SM00034; CLECT; 1.  
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.  
DR PROSITE; PS50041; C-TYPE LECTIN\_2; 1.  
KW Lectin.  
FT DOMAIN 1 129 C-TYPE LECTIN (LONG FORM).  
FT DISULFID 2 13 BY SIMILARITY.  
FT DISULFID 30 127 BY SIMILARITY.  
FT DISULFID 102 119 BY SIMILARITY.  
SQ SEQUENCE 131 AA; 15427 MW; B3569F5BF91F6624 CRC64;  
Query Match 4.3%; Score 6; DB 1; Length 131;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 119 LEKRAE 124  
DB 105 LEKRAE 110  
|||||  
RESULT 31  
RL24\_AERPE STANDARD; PRT; 132 AA.  
AC QYF83;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 18-OCT-2001 (Rel. 40, Last annotation update)  
DE 50S ribosomal protein L24p.  
GN RPL24P OR APE0358.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
OC Desulfurococaceae; Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1;  
RX MEDLINE=99310339; PubMed=10382966;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
RA Jinno K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,  
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1."  
RL DNA Res. 6:83-101(1999).  
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AP000059; BAA79313.1; -  
DR PIR; E72727; E72727.  
DR InterPro; IPR005824; KOW.  
DR InterPro; IPR006646; KOW sub.  
DR InterPro; IPR005756; L26\_euk\_arch.  
DR InterPro; IPR005825; Ribosomal\_L24\_26.  
DR Pfam; PF00467; KOW; 1.  
DR SMART; SM00739; KOW; 1.  
DR TIGRFAMs; TIGR01080; rplX\_A\_E; 1.  
DR PROSITE; PS01108; RIBOSOMAL\_L24; 1.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 132 AA; 15129 MW; 58A82CE320F3FC96 CRC64;  
Query Match 4.3%; Score 6; DB 1; Length 132;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 84 SPELRE 89  
DB 34 SPELRE 39  
|||||  
RESULT 32  
APB\_RABIT STANDARD; PRT; 144 AA.  
AC PL1765;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-AUG-1990 (Rel. 15, Last annotation update)  
DE Apolipoprotein B (Fragment).  
GN APOB.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88335569; PubMed=3419914;  
RX Sudarickov A., Surguchov A.;  
RT "Probe for rabbit apolipoprotein B gene."  
RL Nucleic Acids Res. 16:8187-8187(1988).  
CC -1- FUNCTION: APOLIPOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF  
CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION  
CC SIGNAL FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL  
CC PARTICLES BY THE APOB/E RECEPTOR.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X07480; CA30366.1; -  
DR Plasma; Lipid transport; VLDL; LDL; Chylomicron; Heparin-binding;  
KW Cholesterol metabolism; Glycoprotein; Atherosclerosis.  
FT NON TER 1 1  
FT NON TER 144 144  
SQ SEQUENCE 144 AA; 15664 MW; 5ED9F9D0A9EFA26 CRC64;  
Query Match 4.3%; Score 6; DB 1; Length 144;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 GASMKL 74

```

Db          92 GASKML 97
|||||
RESULT 33
VP8_BPPH6   STANDARD;      PRT;    149 AA.
AC   P07579;
DT   01-APR-1998 (Rel. 07, Created)
DT   01-APR-1998 (Rel. 07, Last sequence update)
DT   01-JAN-1990 (Rel. 13, Last annotation update)
DE   Major nucleocapsid protein (P8 protein).
GN   P8.
OS   Bacteriophage phi-6.
OC   Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
OX   NCBI_TaxID=10879;
RN   [1]
RX   MEDLINE=86144095; PubMed=3754015;
RA   McGraw T., Mindich L., Frangione B.;
RT   "Nucleotide sequence of the small double-stranded RNA segment of
RT   bacteriophage phi 6: novel mechanism of natural translational
RT   control.";
RL   J. Virol. 58:142-151(1986).
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; M12921; AAA32358.1; -.
DR   F01; A23368; VHBPF6.
KW   Nucleocapsid.
SQ   SEQUENCE 149 AA; 16004 MW; ECEED0D195186A929 CRC64;
Query Match      4.3%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred.No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY   59 SAIATAT 64
Db          16 SAIATAT 21
|||||
RESULT 34
HES2_MOUSE
ID   HES2_MOUSE   STANDARD;      PRT;    157 AA.
AC   O54792;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Transcription factor HES-2 (Hairy and enhancer of split 2).
GN   HES2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/J;
RX   MEDLINE=98234545; PubMed=9570950;
RA   Nishimura M., Isaka F., Ishibashi M., Tomita K., Tsuda H.,
RA   Nakaniishi S., Kageyama R.;
RT   "Structure, chromosomal locus, and promoter of mouse HES2 gene, a
RT   homologue of Drosophila hairy and Enhancer of split.";
RL   Genomics 49:69-75(1998).
CC   -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF GENES THAT REQUIRE A BHLH
CC   PROTEIN FOR THEIR TRANSCRIPTION.
CC   -!- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX
CC   WITH A CO-REPRESSOR PROTEIN (GROUCHO).

```

```

CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A
CC   HELIX-INTERRUPTING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),
CC   RATHER THAN THE CANONICAL E-BOX (CANNTG).
CC   -!- DOMAIN: THE CARBOXYL-TERMINAL WRW MOTIF IS A TRANSCRIPTIONAL
CC   REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO, A
CC   TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY
CC   HAIRY-RELATED PROTEINS.
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC   TRANSCRIPTION FACTORS "HAIRY" SUBFAMILY.
CC   -!- SIMILARITY: Contains 1 orange domain.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; AB009967; BAA24091.1; -.
DR   MGD; MGI:1098624; Hes2.
DR   InterPro; IPR001092; HLH basic.
DR   InterPro; IPR003650; Orange.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
DR   SMART; SM00511; ORANGE; 1.
DR   PROSITE; PS00038; HLH 1; 1.
DR   PROSITE; PS00888; HLH 2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein; Repressor.
FT   DNA_BIND 13..26
FT   DOMAIN 27..71
FT   DOMAIN 139..148
FT   DOMAIN 154..157
FT   DOMAIN 157..157
FT   DOMAIN 157..157
SQ   SEQUENCE 157 AA; 17231 MW; 570A0C67F4992EA7 CRC64;
Query Match      4.3%; Score 6; DB 1; Length 157;
Best Local Similarity 100.0%; Pred.No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY   117 PLEKR 122
Db          19 PLEKR 24
|||||
RESULT 35
HES2_RAT
ID   HES2_RAT     STANDARD;      PRT;    157 AA.
AC   F35429;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Transcription factor HES-2 (Hairy and enhancer of split 2).
GN   HES2 OR HES-2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryonic brain;
RX   MEDLINE=93358886; PubMed=8354270;
RA   Ishibashi M., Sasai Y., Nakaniishi S., Kageyama R.;
RT   "Molecular characterization of HES-2, a mammalian helix-loop-helix
RT   factor structurally related to Drosophila hairy and Enhancer of
RT   split.";
RL   Eur. J. Biochem. 215:645-652(1993).
CC   -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF GENES THAT REQUIRE A BHLH
CC   PROTEIN FOR THEIR TRANSCRIPTION.
CC   -!- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX
CC   WITH A CO-REPRESSOR PROTEIN (GROUCHO).
CC   -!- SUBCELLULAR LOCATION: Nuclear.

```

CC -!- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A  
CC HELIX-INTERUPTING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),  
CC RATHER THAN THE CANONICAL E-BOX (CANNTG).  
CC  
CC -!- DOMAIN: THE CARBOXYL-TERMINAL WRPW MOTIF IS A TRANSCRIPTIONAL  
CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO. A  
CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY  
CC HAIRY-RELATED PROTEINS.  
CC  
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.  
CC  
CC -!- SIMILARITY: Contains 1 orange domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D14029; BAA03118.1; -.  
CC PIR; S35037; S35037.  
CC TRANSFAC; T01650; -.  
CC InterPro; IPR001092; HLH basic.  
CC InterPro; IPR003650; Orange.  
CC Pfam; PF00010; HLH; 1.  
CC SMART; SM00353; HLH; 1.  
CC SMART; SM00511; ORANGE; 1.  
CC PROSITE; PS00038; HLH 1; 1.  
CC PROSITE; PS00885; HLH 2; 1.  
CC Transcription regulation; DNA-binding; Nuclear protein; Repressor.  
CC  
CC DNA BIND 13 26  
CC BASIC DOMAIN.  
CC  
CC DOMAIN 27 71  
CC HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
CC  
CC DOMAIN 139 148  
CC PRO-RICH.  
CC  
CC DOMAIN 154 157  
CC WRPW MOTIF (REQUIRED FOR ACTIVITY)  
CC (BY SIMILARITY).  
CC  
CC SEQUENCE 157 AA; 17028 MW; B5D621E814AE0369 CRC64;  
CC  
CC Query Match 4.3%; Score 6; DB 1; Length 157;  
CC Best Local Similarity 100.0%; Pred. No. 92;  
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC  
CC Qy 117 PLEKR 122  
CC |||||  
CC 19 PLEKR 24  
CC  
CC RESULT 36  
CC YRN5\_CAEEL STANDARD; PRT; 160 AA.  
CC AC Q09419;  
CC DT 01-NOV-1995 (Rel. 32, Created)  
CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
CC DT 01-NOV-1997 (Rel. 35, Last annotation update)  
CC DE Hypothetical 16.4 kDa protein R07B1.5 in chromosome X precursor.  
CC GN R07B1.5.  
CC OS Caenorhabditis elegans.  
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
CC Rhabditidae; Peloderinae; Caenorhabditis.  
CC OX NCBI\_TaxID=6239;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN=Bristol N2;  
CC RA Kershaw J.;  
CC RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: SOME, TO C.ELEGANS C30G12.4.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC EMBL; Z48621; CAA89542.1; -.  
CC PIR; T23995; T23995.  
CC WormPep; R07B1.5; CE01631.  
CC InterPro; IPR002601; C6.  
CC Pfam; PF01681; C6; 1.  
CC Hypothetical protein; Signal.  
CC FT SIGNAL 1 18  
CC CHAIN 19 160  
CC SER/THR-RICH.  
CC FT DOMAIN 22 58  
CC SEQUENCE 160 AA; 16413 MW; E9C21E219712A13 CRC64;  
CC  
CC Query Match 4.3%; Score 6; DB 1; Length 160;  
CC Best Local Similarity 100.0%; Pred. No. 93;  
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC  
CC Qy 3 STSTTT 8  
CC |||||  
CC 44 STSTTT 49  
CC  
CC RESULT 37  
CC GTR4\_CANFA  
CC ID GTR4\_CANFA STANDARD; PRT; 162 AA.  
CC AC Q9XST2;  
CC DT 28-FEB-2003 (Rel. 41, Created)  
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
CC DE Solute carrier family 2, facilitated glucose transporter, member 4  
CC (Glucose transporter type 4, insulin-responsive) (fragment).  
CC GN SLC2A4 OR GLUT4.  
CC OS Canis familiaris (Dog).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
CC NCBI\_TaxID=9615;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RC TISSUE=Thyroid;  
CC RX MEDLINE=20422104; PubMed=10964405;  
CC RA Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,  
CC Christophe D.;  
CC RT "A method for the large-scale cloning of nuclear proteins and nuclear  
CC targeting sequences on a functional basis.";  
CC RL Anal. Biochem. 284:231-239(2000).  
CC -!- FUNCTION: Insulin-regulated facilitative glucose transporter.  
CC -!- SUBUNIT: Binds to DAXX (By similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Localizes  
CC primarily to the perinuclear region, undergoing continued  
CC recycling to the plasma membrane where it is rapidly  
CC reinternalized. The dileucine internalization motif is critical  
CC for intracellular sequestration (By similarity).  
CC -!- PTM: Sumoylated (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE  
CC TRANSPORTERS SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AJ389533; CAB46835.1; -.  
CC InterPro; IPR005828; Sub transporter.  
CC InterPro; IPR005829; Sug transporter.  
CC Pfam; PF00083; sugar tr; 1  
CC PRINTS; PR00171; SUGTRANSPORT.  
CC PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
CC PROSITE; PS00217; SUGAR\_TRANSPORT\_2; PARTIAL.  
CC Transmembrane; Sugar transporter; Transport; Multigene family.  
CC NON\_TER 1  
CC FT

FT DOMAIN <1 13 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 14 34 8 (POTENTIAL).  
 FT DOMAIN 35 43 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 44 64 9 (POTENTIAL).  
 FT DOMAIN 65 75 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 76 96 10 (POTENTIAL).  
 FT DOMAIN 97 107 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 108 128 11 (POTENTIAL).  
 FT DOMAIN 129 135 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 136 156 12 (POTENTIAL).  
 FT DOMAIN 157 >162 CYTOPLASMIC (POTENTIAL).  
 FT NON TER 162 162  
 SQ SEQUENCE 162 AA; 17453 MW; 0C58CB23C6AD2BB CRC64;

Query Match 4.3%; Score 6; DB 1; Length 162;  
 Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 LAGMCG 49  
 Db 47 LAGMCG 52

## RESULT 38

B3AR\_MERUN  
 ID B3AR\_MERUN STANDARD; PRT; 167 AA.  
 AC 070432;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Beta-3 adrenergic receptor (Fragment).  
 GN ADRB3.  
 OS Meriones unguiculatus (Mongolian jird) (Mongolian gerbil).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;  
 OC Meriones.  
 CC NCBI\_TaxID=10047;  
 RN [1]

SEQUENCE FROM N.A.  
 RC TISSUE=Brain, Fat, and Vestibular labyrinth;  
 RX MEDLINE=99329129; PubMed=10398761;  
 RA Wangemann P., Liu J., Shimozone M., Scofield M.A.;  
 RT "Beta1-adrenergic receptors but not beta2-adrenergic or vasopressin  
 RT receptors regulate K+ secretion in vestibular dark cells of the inner  
 RT ear.";  
 RL J. Membr. Biol. 170:67-77(1999).  
 CC -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-  
 CC INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G  
 CC PROTEINS. BETA-3 IS INVOLVED IN THE REGULATION OF LIPOLYSIS AND  
 CC THERMOGENESIS.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AF055351; AAC12769.1;  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm1.1;  
 DR PRINTS; PS00237; GPCRHOODPSN  
 DR PROSITE; PS00237; G-PROTEIN\_RECEP\_F1\_1; PARTIAL.  
 DR PROSITE; PS00262; G-PROTEIN\_RECEP\_F1\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Multigene family.  
 FT NON TER 1 1  
 FT DOMAIN <1 25 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 26 47 5 (POTENTIAL).  
 FT DOMAIN 48 114 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 115 136 6 (POTENTIAL).

FT DOMAIN 137 148 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 149 >167 7 (POTENTIAL).  
 FT NON TER 167 167  
 SQ SEQUENCE 167 AA; 18324 MW; CEA78CF68DEEB1C1 CRC64;

Query Match 4.3%; Score 6; DB 1; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 LRALAG 46  
 Db 136 LRALAG 141

## RESULT 39

ILVH\_MYCAV  
 ID ILVH\_MYCAV STANDARD; PRT; 167 AA.  
 AC Q59499;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Acetolactate synthase small subunit (sc.2.2.1.6) (AHAS)  
 DE (Acetohydroxy-acid synthase small subunit) (ALS).  
 GN ILVH OR ILVN.  
 OS Mycobacterium avium.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 CC NCBI\_TaxID=1764;  
 RN [1]

SEQUENCE FROM N.A.  
 RP MEDLINE=97080504; PubMed=8921849;  
 RX Gusberti L., Cantoni R., de Rossi E., Branzoni M., Riccardi G.;  
 RT "Cloning and sequencing of the ilvNC gene cluster from Mycobacterium  
 RT avium.";  
 RL Gene 177:83-85 (1996).  
 CC -!- CATALYTIC ACTIVITY: 2 pyruvate = 2-acetolactate + CO(2).  
 CC -!- PATHWAY: Valine and isoleucine biosynthesis; first step.  
 CC -!- SUBUNIT: DIMER OF LARGE AND SMALL CHAINS (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE ACETOLACTATE SYNTHASE SMALL SUBUNIT  
 CC FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L49392; AAB38427.1;  
 DR InterPro; IPR004789; Acolac\_sm.  
 DR InterPro; IPR002912; ACT.  
 DR Pfam; PF01842; ACT; 1.  
 DR ProDom; PD002844; Acolac\_sm; 1.  
 DR TIGRFAMs; TIGR00119; acolac\_sm; 1.  
 KW Transferase; Branched-chain-amino acid biosynthesis.  
 SQ SEQUENCE 167 AA; 18131 MW; 2F234C6CE9B8DD2D CRC64;

Query Match 4.3%; Score 6; DB 1; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 EALLRV 32  
 Db 135 EALLRV 140

## RESULT 40

ILVH\_MYCTU  
 ID ILVH\_MYCTU STANDARD; PRT; 168 AA.  
 AC O53219;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

Search completed: November 14, 2003, 10:51:04  
Job time : 30 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:48:21 ; Search time 34 Seconds

(without alignments)  
1070.159 Million cell updates/sec

Title: US-10-087-573-2

Perfect score: 141  
Sequence: 1 MESTSTTNFVAENRPTGEE.....RAEYFRHLRLSLKSGQVNRLLI 141

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

- SPTREMBL 23:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_prodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141	100.0	141	5 Q8MNM4	Q8mmn4 babesia can
2	119	84.4	285	5 Q8MQH4	Q8mqh4 babesia can
3	8	5.7	334	10 Q9LHS1	Q9lhs1 arabidopsis
4	8	5.7	434	10 Q9FG07	Q9fg07 arabidopsis
5	8	5.7	513	10 Q9STM8	Q9stm8 arabidopsis
6	8	5.7	534	16 Q8SGU2	Q8sec2 shewanella
7	8	5.7	559	10 Q9SK35	Q9ek35 arabidopsis
8	8	5.7	609	10 Q9FH26	Q9fn26 arabidopsis
9	8	5.7	689	10 Q9FG06	Q9fg06 arabidopsis
10	8	5.7	707	10 Q9STM7	Q9stm7 arabidopsis
11	7	5.0	65	16 Q9BDQ2	Q9bdq2 streptomyce
12	7	5.0	91	5 Q9BNJ8	Q9bnj8 platydesm
13	7	5.0	105	2 Q49812	Q49812 mycobacteri
14	7	5.0	133	16 Q8DF92	Q8df92 vibrio vuln
15	7	5.0	148	2 Q49876	Q49876 mycobacteri
16	7	5.0	159	5 Q9VCT2	Q9vct2 drosophila

17	7	5.0	172	6 Q9BDZ0	Q9bdz0 macaca mula
18	7	5.0	203	5 Q9VH39	Q9vh39 drosophila
19	7	5.0	223	6 Q9SLB5	Q9slb5 ovis aries
20	7	5.0	225	16 Q8FWQ6	Q8fwq6 bruceella su
21	7	5.0	237	16 Q8YBM1	Q8ybm1 bruceella me
22	7	5.0	243	16 Q8B7C4	Q8b7c4 rhizobium l
23	7	5.0	255	5 Q9SSR4	Q9ssr4 drosophila
24	7	5.0	269	2 Q93UL3	Q93ul3 rhizobium g
25	7	5.0	270	2 Q93UK8	Q93uk8 sinorhizobi
26	7	5.0	273	2 Q8RNL4	Q8rnl4 sinorhizobi
27	7	5.0	281	2 Q93UL5	Q93ul5 rhizobium e
28	7	5.0	282	2 Q93UL4	Q93ul4 rhizobium l
29	7	5.0	283	16 Q8DID3	Q8did3 synecococc
30	7	5.0	283	17 Q8U2B4	Q8u2b4 pyrococcus
31	7	5.0	286	2 Q93UK9	Q93uk9 rhizobium e
32	7	5.0	290	2 Q93UL2	Q93ul2 rhizobium g
33	7	5.0	297	12 Q56780	Q56780 european ba
34	7	5.0	308	3 Q9C1K6	Q9cl6 neurospora
35	7	5.0	311	2 Q93UK6	Q93uk6 rhizobium l
36	7	5.0	323	16 Q8XCX9	Q8xcx9 salmonella
37	7	5.0	330	16 Q9I4V9	Q9i4v9 pseudomonas
38	7	5.0	331	16 Q9AD29	Q9ad29 streptomyce
39	7	5.0	332	6 Q8HXX3	Q8hxx3 macaca fasc
40	7	5.0	332	17 Q97YG7	Q97yg7 sulfolobus
41	7	5.0	338	2 Q8RPZ8	Q8rpz8 azotobacter
42	7	5.0	339	5 Q18567	Q18567 caenorhabdi
43	7	5.0	342	17 Q8TKF9	Q8tkf9 methanosaec
44	7	5.0	378	16 Q92DA8	Q92da8 listeria in
45	7	5.0	378	16 Q8Y8J3	Q8y8j3 listeria mo
46	7	5.0	391	6 Q95MM0	Q95mm0 cryptolagus
47	7	5.0	399	2 Q8RJM6	Q8rjm6 vibrio chol
48	7	5.0	409	16 P95182	P95182 mycobacteri
49	7	5.0	409	16 Q53957	Q53957 mycobacteri
50	7	5.0	410	2 Q99Q11	Q99q11 mycobacteri
51	7	5.0	422	17 Q8TVI6	Q8tv16 methanopyru
52	7	5.0	423	16 Q8EJP7	Q8ejp7 shewanella
53	7	5.0	427	5 Q8MP98	Q8mp98 dictyosteli
54	7	5.0	443	2 Q8KLG3	Q8klg3 rhizobium e
55	7	5.0	456	16 Q8XPX8	Q8xpx8 raistonia s
56	7	5.0	473	5 Q94416	Q94416 caenorhabdi
57	7	5.0	475	17 Q8U2Y3	Q8u2y3 pyrococcus
58	7	5.0	490	16 Q8D3L6	Q8d3l6 vibrio vuln
59	7	5.0	491	16 Q9I2I8	Q9i2i8 pseudomonas
60	7	5.0	494	10 Q8RYN6	Q8ryn6 oryza sativ
61	7	5.0	496	5 Q8SWF5	Q8swf5 encephalito
62	7	5.0	499	5 Q966L0	Q966l0 caenorhabdi
63	7	5.0	512	2 Q87636	Q87636 pseudomonas
64	7	5.0	512	5 Q9V3A8	Q9v3a8 drosophila
65	7	5.0	519	16 Q9KXB5	Q9kxb5 streptomyce
66	7	5.0	524	16 Q53294	Q53294 mycobacteri
67	7	5.0	527	16 Q9I3H5	Q9i3h5 pseudomonas
68	7	5.0	538	3 Q43011	Q43011 schizosacch
69	7	5.0	591	2 Q8GNK9	Q8gnk9 mycobacteri
70	7	5.0	614	16 Q8FWB2	Q8fwb2 bruceella su
71	7	5.0	616	5 Q9BN24	Q9bn24 leishmania
72	7	5.0	637	16 Q8D3W6	Q8d3w6 vibrio vuln
73	7	5.0	670	8 Q9G927	Q9g927 ochronas
74	7	5.0	672	13 Q9W7B7	Q9w7b7 brachydanio
75	7	5.0	677	16 Q8PD55	Q8pd55 xanthomonas
76	7	5.0	687	13 Q9IAL0	Q9ial0 brachydanio
77	7	5.0	688	13 Q9W7R7	Q9w7r7 xenopus lae
78	7	5.0	688	13 Q9IBCB	Q9ibcb xenopus lae
79	7	5.0	694	16 Q98DS8	Q98ds8 rhizobium l
80	7	5.0	701	16 Q8PGQ5	Q8pgq5 xanthomonas
81	7	5.0	716	10 Q9SP26	Q9sp26 pisum sativ
82	7	5.0	731	16 Q8ZS53	Q8zss3 anabaena sp
83	7	5.0	747	3 Q8TFX3	Q8tfx3 aspergillus
84	7	5.0	748	16 Q8X6K3	Q8x6k3 escherichia
85	7	5.0	748	16 Q8FEB0	Q8feb0 escherichia
86	7	5.0	753	5 Q44861	Q44861 caenorhabdi
87	7	5.0	773	5 Q8IB77	Q8ib77 plasmodium
88	7	5.0	810	11 Q9ESZ9	Q9esz9 mus musculu
89	7	5.0	841	10 Q64542	Q64542 arabidopsis

90	7	5.0	860	10	Q8LIHO	Q8LIHO oryza sativ	163	6	4.3	112	12	Q91IB3	Q91IB3 influenza a
91	7	5.0	860	11	Q8C9L8	Q8C9L8 mus musculus	164	6	4.3	112	12	Q91IB0	Q91IB0 influenza a
92	7	5.0	869	10	Q9FTV4	Q9FTV4 oryza sativ	165	6	4.3	112	12	Q91IA8	Q91IA8 influenza a
93	7	5.0	925	8	Q9G8S2	Q9G8S2 flammulina	166	6	4.3	113	16	Q8CKG0	Q8CKG0 yersinia pe
94	7	5.0	926	11	Q8R554	Q8R554 mus musculus	167	6	4.3	116	2	Q9RQI9	Q9RQI9 listeria mo
95	7	5.0	964	16	Q8R554	Q8R554 xanthomonas	168	6	4.3	116	16	Q9KMR2	Q9KMR2 vibrio chol
96	7	5.0	1088	10	Q9MBC2	Q9MBC2 hordeum vul	169	6	4.3	116	16	Q9Y9Z9	Q9Y9Z9 aeropyrum p
97	7	5.0	1175	17	Q8T123	Q8T123 methanosarc	170	6	4.3	117	16	Q9ACS9	Q9ACS9 streptomyce
98	7	5.0	1175	17	Q8P333	Q8P333 methanosarc	171	6	4.3	117	16	Q8DI75	Q8DI75 synchococc
99	7	5.0	1184	16	Q83909	Q83909 treponema p	172	6	4.3	117	16	Q8DI75	Q8DI75 synchococc
100	7	5.0	1203	5	Q8T1J7	Q8T1J7 dictyosteli	173	6	4.3	118	5	Q9NJH4	Q9NJH4 trypanosoma
101	7	5.0	1218	5	Q960H7	Q960H7 drosophila	174	6	4.3	119	5	Q9VPT0	Q9VPT0 drosophila
102	7	5.0	1387	5	Q8I0C8	Q8I0C8 drosophila	175	6	4.3	120	16	Q8FI08	Q8FI08 leptospira
103	7	5.0	1392	5	Q9V745	Q9V745 drosophila	176	6	4.3	121	17	Q97B18	Q97B18 thermoplasm
104	7	5.0	1547	5	Q01790	Q01790 caenorhabdi	177	6	4.3	122	4	Q9B7B7	Q9B7B7 homo sapien
105	7	5.0	2075	5	Q8IHR5	Q8IHR5 plasmodium	178	6	4.3	122	4	Q9B7B7	Q9B7B7 homo sapien
106	7	5.0	2422	16	Q8P100	Q8P100 xanthomonas	179	6	4.3	122	16	Q9MTI9	Q9MTI9 oenothera h
107	7	5.0	2833	5	Q9V2P3	Q9V2P3 drosophila	180	6	4.3	122	16	Q8Z599	Q8Z599 yersinia pe
108	7	5.0	2945	5	Q95XG5	Q95XG5 dictyosteli	181	6	4.3	123	12	Q8O842	Q8O842 hantavirus
109	7	5.0	4385	5	Q8I0G2	Q8I0G2 caenorhabdi	182	6	4.3	123	16	Q9PGV5	Q9PGV5 xylella fas
110	7	5.0	4498	5	Q9W2Z3	Q9W2Z3 drosophila	183	6	4.3	123	17	Q96XX0	Q96XX0 sulfolobus
111	6	4.3	29	9	Q9XJNS	Q9XJNS bacterioph	184	6	4.3	126	9	Q942X3	Q942X3 oryza sativ
112	6	4.3	34	2	Q9R5B3	Q9R5B3 streptomyce	185	6	4.3	126	8	Q35408	Q35408 perga conde
113	6	4.3	38	9	Q9XJN1	Q9XJN1 bacterioph	186	6	4.3	128	8	Q34067	Q34067 streptococc
114	6	4.3	45	12	Q68607	Q68607 hepatitis c	187	6	4.3	128	9	Q8O175	Q8O175 streptococc
115	6	4.3	54	5	Q9U7K4	Q9U7K4 plasmodium	188	6	4.3	128	9	Q8O175	Q8O175 streptococc
116	6	4.3	57	5	Q26196	Q26196 pyrococcus	189	6	4.3	128	11	Q9QZ99	Q9QZ99 mus musculu
117	6	4.3	59	5	Q9B7F8	Q9B7F8 plasmodium	190	6	4.3	128	11	Q9QZ99	Q9QZ99 mus musculu
118	6	4.3	59	12	Q9WPH3	Q9WPH3 equine herp	191	6	4.3	129	4	Q9HAI2	Q9HAI2 homo sapien
119	6	4.3	59	16	Q8VJ64	Q8VJ64 mycobacteri	192	6	4.3	130	10	Q8GWM5	Q8GWM5 arabidopsis
120	6	4.3	60	12	Q96699	Q96699 canine herp	193	6	4.3	130	12	Q8JQF5	Q8JQF5 equine herp
121	6	4.3	60	12	Q96744	Q96744 feline herp	194	6	4.3	130	15	Q92S97	Q92S97 rhizobium m
122	6	4.3	60	12	Q8BAH2	Q8BAH2 lung-eye-tr	195	6	4.3	131	17	Q9HSY0	Q9HSY0 halobacteri
123	6	4.3	62	16	Q8X4E7	Q8X4E7 escherichia	196	6	4.3	131	17	Q8Z2P3	Q8Z2P3 pyrobaculum
124	6	4.3	63	6	Q8MJ53	Q8MJ53 equus cabal	197	6	4.3	132	4	Q9H8K5	Q9H8K5 homo sapien
125	6	4.3	63	12	Q9QC11	Q9QC11 bottlenose	198	6	4.3	134	17	Q97AF3	Q97AF3 thermoplasm
126	6	4.3	67	2	Q93PX7	Q93PX7 pseudomonas	199	6	4.3	135	5	Q8MZ97	Q8MZ97 drosophila
127	6	4.3	67	16	Q9KOR1	Q9KOR1 neisseria m	200	6	4.3	135	5	Q9VG36	Q9VG36 drosophila
128	6	4.3	71	2	Q93M44	Q93M44 bordetella	201	6	4.3	135	16	Q9ABQ0	Q9ABQ0 caulobacter
129	6	4.3	73	2	Q8KN06	Q8KN06 bacillus th	202	6	4.3	137	16	Q9ABP9	Q9ABP9 caulobacter
130	6	4.3	75	16	Q928Y4	Q928Y4 listeria in	203	6	4.3	141	4	Q9Y425	Q9Y425 homo sapien
131	6	4.3	76	4	Q96EN4	Q96EN4 homo sapien	204	6	4.3	141	9	Q94M65	Q94M65 streptococc
132	6	4.3	77	4	Q91945	Q91945 caenorhabdi	205	6	4.3	142	11	Q920V9	Q920V9 mus musculu
133	6	4.3	78	16	Q93M94	Q93M94 clostridium	206	6	4.3	143	11	Q92D95	Q92D95 mus musculu
134	6	4.3	80	16	Q8VJ99	Q8VJ99 mycobacteri	207	6	4.3	143	16	Q9A540	Q9A540 caulobacter
135	6	4.3	82	10	Q8O943	Q8O943 arabidopsis	208	6	4.3	144	16	Q9AWC1	Q9AWC1 agrobacteri
136	6	4.3	86	16	Q9AGE8	Q9AGE8 caulobacter	209	6	4.3	145	5	Q61351	Q61351 drosophila
137	6	4.3	89	15	Q9KUX5	Q9KUX5 vibrio chol	210	6	4.3	145	5	Q9Y122	Q9Y122 hydra magni
138	6	4.3	90	15	Q66735	Q66735 equine infe	211	6	4.3	145	9	Q38182	Q38182 lactococcus
139	6	4.3	91	5	Q9ENW6	Q9ENW6 hiltonius s	212	6	4.3	145	9	Q9B014	Q9B014 bacterioph
140	6	4.3	91	5	Q9ENW6	Q9ENW6 hiltonius s	213	6	4.3	145	16	Q9CGT2	Q9CGT2 lactococcus
141	6	4.3	95	10	Q9LDM6	Q9LDM6 oryza sativ	214	6	4.3	146	3	Q9P492	Q9P492 podospira a
142	6	4.3	96	15	Q8USX3	Q8USX3 human immun	215	6	4.3	146	3	Q9P492	Q9P492 podospira a
143	6	4.3	97	16	Q8NQS2	Q8NQS2 corynebacte	216	6	4.3	147	5	Q8T176	Q8T176 bacterioph
144	6	4.3	98	16	Q8RFP8	Q8RFP8 fusobacteri	217	6	4.3	148	5	Q9GTU0	Q9GTU0 dictyosteli
145	6	4.3	100	12	Q91LLA	Q91LLA white spot	218	6	4.3	148	10	Q9LXH1	Q9LXH1 arabidopsis
146	6	4.3	101	10	Q8VYX7	Q8VYX7 pisum sativ	219	6	4.3	148	16	Q820Q3	Q820Q3 anabaena sp
147	6	4.3	101	17	Q9Y8V3	Q9Y8V3 aeropyrum p	220	6	4.3	149	3	Q8TFM2	Q8TFM2 aspergillus
148	6	4.3	102	16	Q8UI56	Q8UI56 agrobacteri	221	6	4.3	149	6	Q95JR2	Q95JR2 macaca fasc
149	6	4.3	103	4	Q9P1G0	Q9P1G0 homo sapien	222	6	4.3	149	9	Q9XJN9	Q9XJN9 bacterioph
150	6	4.3	103	17	Q9V032	Q9V032 pyrococcus	223	6	4.3	151	2	Q9RQK4	Q9RQK4 mycoplasma
151	6	4.3	104	2	Q9EY10	Q9EY10 streptomyce	224	6	4.3	151	5	Q95VQ0	Q95VQ0 microtheca y
152	6	4.3	104	16	Q9WYV7	Q9WYV7 thermotoga	225	6	4.3	151	8	Q9G6Z2	Q9G6Z2 microtheca y
153	6	4.3	104	16	Q9CD46	Q9CD46 mycobacteri	226	6	4.3	151	10	Q9FFY3	Q9FFY3 arabidopsis
154	6	4.3	105	2	Q9AEZ3	Q9AEZ3 frankia sp.	227	6	4.3	151	13	Q90XG5	Q90XG5 prototenus
155	6	4.3	105	5	Q8SXZ5	Q8SXZ5 drosophila	228	6	4.3	152	4	Q14713	Q14713 homo sapien
156	6	4.3	107	15	Q8USY0	Q8USY0 human immun	229	6	4.3	152	11	Q8C2P0	Q8C2P0 mus musculu
157	6	4.3	108	16	Q8DLK4	Q8DLK4 synchococc	230	6	4.3	153	16	Q8E390	Q8E390 streptococc
158	6	4.3	111	3	Q04293	Q04293 pichia angu	231	6	4.3	153	16	Q8DXM0	Q8DXM0 oryza sativ
159	6	4.3	111	5	Q9NFG9	Q9NFG9 plasmodium	232	6	4.3	154	10	Q9XEU8	Q9XEU8 hepatitis c
160	6	4.3	112	12	Q91IB1	Q91IB1 influenza a	233	6	4.3	154	12	Q917C2	Q917C2 vibrio chol
161	6	4.3	112	12	Q91IB2	Q91IB2 influenza a	234	6	4.3	154	16	Q9XPT5	Q9XPT5 streptomyce
162	6	4.3	112	12	Q91IB2	Q91IB2 influenza a	235	6	4.3	154	16	Q86495	Q86495 streptomyce



236	6	4.3	155	2	Q9WWT8	Q9WVt8 pseudomonas	309	6	4.3	188	11	Q04365	Q04365 mus musculus
237	6	4.3	156	2	Q9FBN3	Q9FBN3 carboxydoth	310	6	4.3	188	13	Q90350	Q90350 coturnix co
238	6	4.3	156	4	Q9FLR6	Q961r6 homo sapien	311	6	4.3	188	16	Q83722	Q83722 treponema p
239	6	4.3	156	12	Q66089	Q66089 canine heip	312	6	4.3	188	16	Q8CR07	Q8CR07 staphylococ
240	6	4.3	157	2	Q93R95	Q93R95 thermus the	313	6	4.3	189	16	Q8NPy3	Q8NPy3 corynebacte
241	6	4.3	157	11	Q8BKAE	Q8bka6 mus musculus	314	6	4.3	190	2	Q8KTW0	Q8ktw0 escherichia
242	6	4.3	157	2	Q9F018	Q9f018 streptococc	315	6	4.3	190	16	Q8DIY7	Q8diy7 synchococc
243	6	4.3	158	10	Q8QW99	Q8qwr-9 arabidopsis	316	6	4.3	191	2	Q9RFS2	Q9rf2 vibrio vuln
244	6	4.3	158	16	Q91206	Q91206 pseudomonas	317	6	4.3	192	2	Q8KLY5	Q8kly5 pseudomonas
245	6	4.3	158	16	Q97SD8	Q97sd8 streptococc	318	6	4.3	192	2	Q8KLY5	Q8kly5 pseudomonas
246	6	4.3	159	16	Q92NG2	Q92ng2 staphylococ	319	6	4.3	192	2	Q8KLY5	Q8kly5 pseudomonas
247	6	4.3	159	16	Q8CPV2	Q8cpv2 staphylococ	320	6	4.3	192	2	Q8KLY5	Q8kly5 pseudomonas
248	6	4.3	160	5	Q81TA7	Q8ita7 aequipecten	321	6	4.3	192	5	Q8WPV7	Q8wpv7 plasmodium
249	6	4.3	160	16	Q9X002	Q9x002 thermotoga	322	6	4.3	192	16	Q8XUS3	Q8xus3 ralstonia s
250	6	4.3	160	16	Q8FVW3	Q8fmv3 corynebacte	323	6	4.3	193	16	Q8XUS3	Q8xus3 ralstonia s
251	6	4.3	161	11	Q9QZG0	Q9qzg0 rattus norv	324	6	4.3	193	16	Q8XUS3	Q8xus3 ralstonia s
252	6	4.3	161	11	Q8BTI3	Q8bti3 mus musculus	325	6	4.3	193	16	Q8XUS3	Q8xus3 ralstonia s
253	6	4.3	162	16	Q8F4U1	Q8f4u1 leptospira	326	6	4.3	193	16	Q8XUS3	Q8xus3 ralstonia s
254	6	4.3	162	17	Q8F4U1	Q8f4u1 leptospira	327	6	4.3	193	16	Q8XUS3	Q8xus3 ralstonia s
255	6	4.3	163	2	Q8REV3	Q8rev3 variovorax	328	6	4.3	195	2	Q8VQV3	Q8vqv3 rhodococcus
256	6	4.3	163	5	Q8WQ90	Q8mq90 caenorhabdi	329	6	4.3	195	2	Q8VQV3	Q8vqv3 rhodococcus
257	6	4.3	164	16	Q9Z582	Q9z582 streptomyce	330	6	4.3	195	2	Q8VQV3	Q8vqv3 rhodococcus
258	6	4.3	164	17	Q9ZUF3	Q9zvf3 pyrococcus	331	6	4.3	195	2	Q8VQV3	Q8vqv3 rhodococcus
259	6	4.3	165	4	Q96PC6	Q96pc6 homo sapien	332	6	4.3	195	2	Q8VQV3	Q8vqv3 rhodococcus
260	6	4.3	165	16	Q9RS29	Q9rs29 deinococcus	333	6	4.3	195	2	Q8VQV3	Q8vqv3 rhodococcus
261	6	4.3	166	5	Q18971	Q18971 caenorhabdi	334	6	4.3	196	2	Q9F5B7	Q9f5b7 agrobacteri
262	6	4.3	166	10	Q8RZ96	Q8rz96 oryza sativ	335	6	4.3	196	2	Q9F5B7	Q9f5b7 agrobacteri
263	6	4.3	166	16	Q8DR04	Q8dr04 streptococc	336	6	4.3	196	2	Q9F5B7	Q9f5b7 agrobacteri
264	6	4.3	167	1	Q9VZ23	Q9vz23 methanoroco	337	6	4.3	197	16	Q8EAP5	Q8ead5 shewanella
265	6	4.3	167	2	Q9X7I3	Q9x7i3 prochloroco	338	6	4.3	197	16	Q8EAP5	Q8ead5 shewanella
266	6	4.3	169	2	Q8KRW5	Q8krs5 rhizobium l	339	6	4.3	198	5	Q9VPW6	Q9vpw6 drosophila
267	6	4.3	169	11	Q9D5G1	Q9d5g1 mus musculus	340	6	4.3	198	5	Q9VPW6	Q9vpw6 drosophila
268	6	4.3	170	12	Q9J3M1	Q9j3m1 human rotav	341	6	4.3	198	10	Q04118	Q04118 pisum sativ
269	6	4.3	171	10	Q8LKF4	Q8lkf4 malus domes	342	6	4.3	198	10	Q04118	Q04118 pisum sativ
270	6	4.3	171	12	Q91IR8	Q91ir8 hepatitis c	343	6	4.3	198	10	Q04118	Q04118 pisum sativ
271	6	4.3	171	12	Q91IR8	Q91ir8 hepatitis c	344	6	4.3	198	10	Q04118	Q04118 pisum sativ
272	6	4.3	171	12	Q91IR8	Q91ir8 hepatitis c	345	6	4.3	198	10	Q04118	Q04118 pisum sativ
273	6	4.3	171	12	Q91IR8	Q91ir8 hepatitis c	346	6	4.3	198	10	Q04118	Q04118 pisum sativ
274	6	4.3	171	12	Q91IR8	Q91ir8 hepatitis c	347	6	4.3	198	10	Q04118	Q04118 pisum sativ
275	6	4.3	171	12	Q91IR8	Q91ir8 hepatitis c	348	6	4.3	198	10	Q04118	Q04118 pisum sativ
276	6	4.3	172	16	Q9A3B5	Q9a3b5 caulobacter	349	6	4.3	199	17	Q9HPM6	Q9hpm6 halobacteri
277	6	4.3	172	16	Q9A3B5	Q9a3b5 caulobacter	350	6	4.3	199	17	Q9HPM6	Q9hpm6 halobacteri
278	6	4.3	173	5	Q8MNI7	Q8mni7 dictyosteli	351	6	4.3	200	12	Q9YTL1	Q9ytl1 ateline her
279	6	4.3	174	10	Q22084	Q22084 petunia hyb	352	6	4.3	201	4	Q9UF72	Q9uf72 homo sapien
280	6	4.3	176	16	Q9AB29	Q9ab29 caulobacter	353	6	4.3	202	10	Q8LWX9	Q8lmx9 oryza sativ
281	6	4.3	177	16	Q8FA24	Q8fa24 xanthomonas	354	6	4.3	203	16	Q91614	Q91614 pseudomonas
282	6	4.3	178	2	Q33538	Q33538 rhodococcus	355	6	4.3	203	16	Q91614	Q91614 pseudomonas
283	6	4.3	178	10	Q9SVD6	Q9svd6 arabidopsis	356	6	4.3	203	16	Q91614	Q91614 pseudomonas
284	6	4.3	178	16	P74585	P74585 synchocyst	357	6	4.3	203	16	Q91614	Q91614 pseudomonas
285	6	4.3	179	16	Q8EX22	Q8ex22 leptospira	358	6	4.3	203	16	Q91614	Q91614 pseudomonas
286	6	4.3	181	4	Q9H674	Q9h674 homo sapien	359	6	4.3	203	16	Q91614	Q91614 pseudomonas
287	6	4.3	181	4	Q9BUC7	Q9buc7 homo sapien	360	6	4.3	203	16	Q91614	Q91614 pseudomonas
288	6	4.3	181	11	Q8K249	Q8k249 mus musculus	361	6	4.3	203	16	Q91614	Q91614 pseudomonas
289	6	4.3	181	11	Q8BGR0	Q8bgr0 mus musculus	362	6	4.3	203	16	Q91614	Q91614 pseudomonas
290	6	4.3	181	16	Q8F915	Q8f915 leptospira	363	6	4.3	203	16	Q91614	Q91614 pseudomonas
291	6	4.3	182	5	Q9XZV0	Q9xzv0 leishmania	364	6	4.3	203	16	Q91614	Q91614 pseudomonas
292	6	4.3	182	8	Q95H21	Q95h21 octoblephar	365	6	4.3	203	16	Q91614	Q91614 pseudomonas
293	6	4.3	182	8	Q95H22	Q95h22 octoblephar	366	6	4.3	203	16	Q91614	Q91614 pseudomonas
294	6	4.3	182	13	Q8QCC6	Q8qcc6 xenopus lae	367	6	4.3	203	16	Q91614	Q91614 pseudomonas
295	6	4.3	182	16	Q8EKP8	Q8ekp8 shewanella	368	6	4.3	203	16	Q91614	Q91614 pseudomonas
296	6	4.3	182	17	Q26796	Q26796 methanobact	369	6	4.3	203	16	Q91614	Q91614 pseudomonas
297	6	4.3	183	2	Q69302	Q69302 campylobact	370	6	4.3	203	16	Q91614	Q91614 pseudomonas
298	6	4.3	183	16	Q9PNM3	Q9pnm3 campylobact	371	6	4.3	203	16	Q91614	Q91614 pseudomonas
299	6	4.3	183	16	Q911Y7	Q911y7 pseudomonas	372	6	4.3	203	16	Q91614	Q91614 pseudomonas
300	6	4.3	184	16	Q8ERM9	Q8erm9 oceanobacil	373	6	4.3	203	16	Q91614	Q91614 pseudomonas
301	6	4.3	184	17	Q8TLV9	Q8tlv9 methanosarc	374	6	4.3	203	16	Q91614	Q91614 pseudomonas
302	6	4.3	186	10	Q9L7D2	Q9ld2 arabidopsis	375	6	4.3	203	16	Q91614	Q91614 pseudomonas
303	6	4.3	186	17	Q8PSA3	Q8psa3 methanosarc	376	6	4.3	203	16	Q91614	Q91614 pseudomonas
304	6	4.3	187	4	Q96DT9	Q96dt9 homo sapien	377	6	4.3	203	16	Q91614	Q91614 pseudomonas
305	6	4.3	187	5	Q8IRM1	Q8irm1 drosophila	378	6	4.3	203	16	Q91614	Q91614 pseudomonas
306	6	4.3	187	11	Q9WUZ5	Q9wuz5 mus musculus	379	6	4.3	203	16	Q91614	Q91614 pseudomonas
307	6	4.3	187	16	Q8X3S7	Q8x3s7 escherichia	380	6	4.3	203	16	Q91614	Q91614 pseudomonas
308	6	4.3	188	4	Q96T59	Q96t59 homo sapien	381	6	4.3	203	16	Q91614	Q91614 pseudomonas

382	6	4.3	218	10	Q24123	O24123 nicotiana s	455	6	4.3	245	16	Q9L170	Q9L170 streptomyce
383	6	4.3	218	11	Q9CXR6	Q9cxr6 mus musculus	456	6	4.3	246	10	Q8LN31	Q8ln31 cryza sativ
384	6	4.3	218	11	Q9ZEB2	Q9zeb2 rickettsia	457	6	4.3	246	16	Q987P6	Q987p6 rhizobium l
385	6	4.3	218	16	Q9FGS5	Q9fgs5 escherichia	458	6	4.3	246	16	Q8XGK1	Q8xgk1 salmonella
386	6	4.3	218	16	Q9D0C3	Q9d0c3 yersinia pe	459	6	4.3	246	16	Q8P4F1	Q8p4f1 xanthomonas
387	6	4.3	219	4	Q96PJ4	Q96pj4 homo sapien	460	6	4.3	247	17	Q8TVU6	Q8tvu6 methanopyru
388	6	4.3	219	16	Q9HZA9	Q9hza9 pseudomonas	461	6	4.3	248	4	Q86J10	Q86j10 homo sapien
389	6	4.3	221	4	Q9HCC6	Q9hcc6 homo sapien	462	6	4.3	248	12	Q919H9	Q919h9 culex nigri
390	6	4.3	221	5	Q95WZ5	Q95wz5 ixodes scap	463	6	4.3	249	2	Q93H55	Q93h55 streptomyce
391	6	4.3	221	10	Q8H3V7	Q8h3v7 cryza sativ	464	6	4.3	249	2	Q50444	Q50444 mycobacteri
392	6	4.3	221	16	Q8XFR1	Q8xfr1 salmonella	465	6	4.3	250	4	Q9NSG9	Q9nsg9 homo sapien
393	6	4.3	221	16	Q8CJX3	Q8cjx3 streptomyce	466	6	4.3	250	16	Q92RJ5	Q92rj5 rhizobium m
394	6	4.3	222	8	Q9WG12	Q9wg12 kongobathia	467	6	4.3	250	17	Q82V13	Q82v13 pyrobaculum
395	6	4.3	222	2	Q68496	Q68496 legionella	468	6	4.3	251	2	Q9FCX3	Q9fcx3 clostridium
396	6	4.3	223	10	Q8ZR33	Q8zrr3 cryza sativ	469	6	4.3	251	9	Q64113	Q64113 bacterioph
397	6	4.3	223	2	Q93X98	Q93x98 antirrhinum	470	6	4.3	251	16	Q34449	Q34449 bacillus su
398	6	4.3	224	8	Q8M834	Q8m834 microstomum	471	6	4.3	252	10	Q9S2Z6	Q9s2z6 arabidopsis
399	6	4.3	225	4	Q96PJ1	Q96pj1 homo sapien	472	6	4.3	252	10	Q8LIX6	Q8lix6 cryza sativ
400	6	4.3	225	6	Q9BFM2	Q9bfm2 tapirus ind	473	6	4.3	252	17	Q9Y984	Q9y984 aeropyrum p
401	6	4.3	225	6	Q9BFM4	Q9bfm4 equus cabal	474	6	4.3	253	10	Q8LHM5	Q8lhm5 cryza sativ
402	6	4.3	225	6	Q98FM3	Q98fm3 ceratotheri	475	6	4.3	253	10	Q40599	Q40599 petunia hyb
403	6	4.3	225	16	Q9PIL2	Q9pil2 xanthomonas	476	6	4.3	253	16	Q8Y5F7	Q8y5f7 listeria in
404	6	4.3	225	16	Q8P793	Q8p793 xanthomonas	477	6	4.3	253	16	Q8P6C1	Q8p6c1 xanthomonas
405	6	4.3	225	17	Q96XX5	Q96xx5 sulfolobus	478	6	4.3	253	16	Q9FC93	Q9fc93 streptomyce
406	6	4.3	225	12	Q66123	Q66123 cowpea mott	479	6	4.3	253	17	Q8TRU6	Q8tru6 methanosarc
407	6	4.3	226	12	Q82WB2	Q82wb2 pyrobaculum	480	6	4.3	253	17	Q8PV49	Q8pv49 methanosarc
408	6	4.3	226	16	Q928Z3	Q928z3 chlamydia p	481	6	4.3	254	5	Q5M012	Q5m012 drosophila
409	6	4.3	226	16	Q8D8J1	Q8d8j1 vibrio vuln	482	6	4.3	254	5	Q8VR78	Q8vr78 escherichia
410	6	4.3	226	17	Q26341	Q26341 methanobact	483	6	4.3	255	2	Q8R7Z4	Q8r7z4 arabidopsis
411	6	4.3	227	16	Q9C4F8	Q9c4f8 lactococcus	484	6	4.3	255	10	Q9S2Z4	Q9s2z4 streptomyce
412	6	4.3	228	2	Q9EUT0	Q9eut0 staphylococ	485	6	4.3	255	16	Q8CNG6	Q8cng6 escherichia
413	6	4.3	228	10	Q94CV1	Q94cv1 cryza sativ	486	6	4.3	255	16	Q8CNG6	Q8cng6 escherichia
414	6	4.3	228	16	Q99U36	Q99u36 staphylococ	487	6	4.3	255	17	Q8TZ29	Q8tzz9 pyrococcus
415	6	4.3	229	10	Q94F05	Q94f05 arabidopsis	488	6	4.3	255	13	Q9PUY7	Q9pu7 osteoglossu
416	6	4.3	229	10	Q8H0X7	Q8h0x7 arabidopsis	489	6	4.3	256	16	Q8CPH6	Q8cp6 staphylococ
417	6	4.3	229	16	Q9PDK5	Q9pdk5 xylella fas	490	6	4.3	257	5	Q97250	Q97250 plasmodium
418	6	4.3	230	10	Q9FP31	Q9fp31 cryza sativ	491	6	4.3	257	16	Q92N06	Q92n06 rhizobium m
419	6	4.3	230	16	Q8EEH9	Q8eeh9 shewanella	492	6	4.3	258	4	Q9BU20	Q9bu20 homo sapien
420	6	4.3	231	10	Q8H555	Q8h555 cryza sativ	493	6	4.3	258	11	Q9QXV0	Q9qxv0 mus musculu
421	6	4.3	231	16	Q8NQD2	Q8ngd2 corynebacte	494	6	4.3	258	11	Q91W26	Q91w26 mus musculu
422	6	4.3	232	2	Q8KPR1	Q8kpr1 synchocococ	495	6	4.3	258	11	Q31094	Q31094 streptomyce
423	6	4.3	232	2	Q8KQ99	Q8kq99 rhizobium l	496	6	4.3	259	2	Q9K7D6	Q9k7d6 bacillus ha
424	6	4.3	232	10	Q9FVW9	Q9fvw9 arabidopsis	497	6	4.3	259	16	Q9UHG2	Q9uhg2 homo sapien
425	6	4.3	232	16	Q8FCX6	Q8fcx6 brucella me	498	6	4.3	260	4	Q9UHG2	Q9uhg2 rattus norv
426	6	4.3	232	16	Q8XRJ0	Q8xjr0 ralstonia s	499	6	4.3	260	11	Q9QXU9	Q9qxu9 corynebacte
427	6	4.3	233	17	Q97U09	Q97u09 sulfolobus	500	6	4.3	260	16	Q8NN19	Q8nn19 arabidopsis
428	6	4.3	234	2	Q85452	Q85452 rhizobium l	501	6	4.3	261	10	Q9LFD5	Q9lpd5 arabidopsis
429	6	4.3	234	16	Q8FVF1	Q8fvf1 brucella su	502	6	4.3	263	4	Q9BRU3	Q9br3 homo sapien
430	6	4.3	235	2	Q45235	Q45235 bradyrhizob	503	6	4.3	263	16	Q8XTA5	Q8xta5 ralstonia s
431	6	4.3	235	11	Q63349	Q63349 rattus norv	504	6	4.3	264	4	Q96HV5	Q96hv5 homo sapien
432	6	4.3	235	16	Q67273	Q67273 aquifex ae	505	6	4.3	264	4	Q8WX40	Q8wx40 homo sapien
433	6	4.3	235	16	Q927T6	Q927t6 chlamydia p	506	6	4.3	264	5	Q8T2L7	Q8c2l7 dictyosteli
434	6	4.3	235	16	Q8FHC0	Q8fhc0 escherichia	507	6	4.3	264	11	Q9D8U2	Q9d8u2 mus musculu
435	6	4.3	236	10	Q8S5W2	Q8s5w2 cryza sativ	508	6	4.3	265	2	Q46240	Q46240 clostridium
436	6	4.3	236	16	Q8YMB7	Q8ymb7 anabaena sp	509	6	4.3	265	4	Q43388	Q43388 homo sapien
437	6	4.3	236	17	Q27708	Q27708 methanobact	510	6	4.3	265	17	Q26265	Q26265 methanobact
438	6	4.3	237	5	Q9U2K4	Q9u2k4 caenorhabdi	511	6	4.3	266	16	Q98P77	Q98p77 rhizobium l
439	6	4.3	237	16	Q8UI25	Q8ui25 agrobacteri	512	6	4.3	267	2	Q54513	Q54513 yersinia en
440	6	4.3	237	16	Q9RUJ3	Q9ruj3 deinococcus	513	6	4.3	267	12	Q99H25	Q99h25 heliocovert
441	6	4.3	238	16	Q8EMR9	Q8emr9 oceanobacil	514	6	4.3	267	16	Q56949	Q56949 yersinia pe
442	6	4.3	239	16	Q9RS43	Q9rs43 deinococcus	515	6	4.3	267	16	Q8U9N5	Q8u9n5 agrobacteri
443	6	4.3	239	16	Q9F172	Q9f172 vibrio vuln	516	6	4.3	269	2	Q9X610	Q9x610 salmonella
444	6	4.3	239	16	Q8CV66	Q8cv66 oceanobacil	517	6	4.3	269	5	Q8ITC0	Q8itc0 aequitecten
445	6	4.3	240	16	Q8X9X0	Q8x9x0 escherichia	518	6	4.3	269	16	Q92W32	Q92w32 rhizobium m
446	6	4.3	240	17	Q97VF7	Q97vf7 sulfolobus	519	6	4.3	270	4	Q9BVX0	Q9bvx0 homo sapien
447	6	4.3	241	2	Q8RLB6	Q8rlb6 comamonas a	520	6	4.3	270	4	Q96PJ2	Q96pj2 homo sapien
448	6	4.3	241	3	Q01110	Q01110 kluyveromyc	521	6	4.3	270	10	Q9FTN3	Q9ftn3 cryza sativ
449	6	4.3	241	16	Q92R45	Q92r45 rhizobium m	522	6	4.3	270	16	Q98D50	Q98d50 rhizobium l
450	6	4.3	243	16	Q98E46	Q98e46 rhizobium l	523	6	4.3	271	16	Q8RE14	Q8re14 fusbacteri
451	6	4.3	243	16	Q50153	Q50153 mycobacteri	524	6	4.3	271	2	Q929R2	Q929r2 bacillus ha
452	6	4.3	244	2	Q47880	Q47880 escherichia	525	6	4.3	271	2	Q93UK7	Q93uk7 rhizobium g
453	6	4.3	244	4	Q9BR78	Q9br78 homo sapien	526	6	4.3	271	16	Q8P3B5	Q8p3b5 xanthomonas
454	6	4.3	244	16	Q926S7	Q926s7 chlamydia p	527	6	4.3	271	16	Q8E8D9	Q8e8d9 shewanella

528	6	4.3	272	17	Q8T223	Q8T23 methanopyru	601	6	4.3	294	16	Q92JV1	Q92jv1 rhizobium m
529	6	4.3	275	2	Q8MV9	Q8mv9 bacillus an	602	6	4.3	295	13	Q9DE14	Q9dei4 anser anser
530	6	4.3	275	5	Q9U3J0	Q9u3j0 caenorhabdi	603	6	4.3	296	2	Q9F8S8	Q9f8s8 streptomyc
531	6	4.3	275	16	Q92YR7	Q92yr7 listeria in	604	6	4.3	296	2	Q8GHA7	Q8gha7 streptomyc
532	6	4.3	275	16	Q8Y478	Q8y478 listeria mo	605	6	4.3	296	10	Q9FW94	Q9fw94 oryza sativ
533	6	4.3	275	16	Q8P2P8	Q8p2p8 streptococ	606	6	4.3	296	16	Q8FJ86	Q8fj86 xanthomonas
534	6	4.3	275	16	Q9ALJ3	Q9alj3 streptococ	607	6	4.3	296	16	Q8P6I7	Q8p6i7 xanthomonas
535	6	4.3	276	2	Q8KKY4	Q8kky4 rhizobium e	608	6	4.3	297	16	Q8Z894	Q8z894 chlamydia p
536	6	4.3	276	16	Q51675	Q51675 borrelia bu	609	6	4.3	297	16	Q8PH15	Q8ph15 xanthomonas
537	6	4.3	277	2	Q8RST7	Q8rst7 erwinia chr	610	6	4.3	298	16	Q8Y0Y7	Q8y0y7 ralstonia s
538	6	4.3	277	4	Q8WH1	Q8wh1 homo sapien	611	6	4.3	300	3	Q9C2R0	Q9c2r0 neurospora
539	6	4.3	277	16	Q98HE1	Q98he1 rhizobium l	612	6	4.3	300	10	Q9AVG1	Q9avgl oryza sativ
540	6	4.3	278	2	Q9ETR2	Q9etr2 corynebacte	613	6	4.3	300	16	Q9AJW1	Q9ajw1 caulobacter
541	6	4.3	278	16	Q8FS37	Q8fs37 corynebacte	614	6	4.3	300	17	Q9HCH2	Q9hch2 halobacteri
542	6	4.3	279	10	Q94CV4	Q94cv4 oryza sativ	615	6	4.3	301	2	Q93B86	Q93b86 mycobacteri
543	6	4.3	280	10	Q9MAL1	Q9mal1 arabidopsis	616	6	4.3	302	13	Q9DDC5	Q9ddc5 meleagris g
544	6	4.3	280	16	Q98AC7	Q98ac7 rhizobium l	617	6	4.3	302	13	Q9DDB7	Q9ddb7 struthio ca
545	6	4.3	281	1	Q8J308	Q8j308 pyrococcus	618	6	4.3	302	16	Q92UE2	Q92ue2 rhizobium m
546	6	4.3	281	10	Q9J340	Q9j340 arabidopsis	619	6	4.3	302	16	Q8XPR4	Q8xpr4 ralstonia s
547	6	4.3	281	16	Q8ZS03	Q8zs03 salmonella	620	6	4.3	304	9	Q9ZXK8	Q9zxk8 bacterioph
548	6	4.3	281	16	Q8Z9P6	Q8z9p6 salmonella	621	6	4.3	304	16	Q8X544	Q8x544 escherichia
549	6	4.3	281	16	Q8EBC5	Q8ebc5 shewanella	622	6	4.3	304	16	Q8CS13	Q8cs13 staphylococ
550	6	4.3	282	5	Q8WP37	Q8wp37 dendronept	623	6	4.3	305	2	Q45818	Q45818 chloroflexu
551	6	4.3	282	10	Q9FL72	Q9fl72 arabidopsis	624	6	4.3	305	2	Q87967	Q87967 streptomyc
552	6	4.3	282	10	Q8LEA8	Q8lea8 arabidopsis	625	6	4.3	305	16	Q98D14	Q98d14 rhizobium l
553	6	4.3	282	13	Q9DDB6	Q9ddb6 sita europ	626	6	4.3	305	16	Q986U5	Q986u5 rhizobium l
554	6	4.3	282	16	Q9AA63	Q9aa63 caulobacter	627	6	4.3	306	16	Q9U4D4	Q9u4d4 xanthomonas
555	6	4.3	283	3	Q9J337	Q9j337 saccharomyc	628	6	4.3	306	16	Q8PLD8	Q8pld8 xanthomonas
556	6	4.3	283	5	Q9VIP7	Q9vip7 drosophila	629	6	4.3	306	17	Q96ZP7	Q96zp7 sulfolobus
557	6	4.3	283	5	Q8KSF4	Q8ksf4 drosophila	630	6	4.3	306	17	Q8ZX54	Q8zxs4 pyrobaculum
558	6	4.3	283	16	Q988E7	Q988e7 rhizobium l	631	6	4.3	307	2	Q9F186	Q9f186 alcaligenes
559	6	4.3	283	17	Q8ZWS1	Q8zws1 pyrobaculum	632	6	4.3	307	2	Q8N4L8	Q8n4l8 homo sapien
560	6	4.3	285	16	Q92ZG2	Q92zg2 rhizobium m	633	6	4.3	307	4	Q9NFK9	Q9nuk9 homo sapien
561	6	4.3	285	17	Q96Y31	Q96y31 methanobact	634	6	4.3	307	4	Q96FI5	Q96fi5 homo sapien
562	6	4.3	286	16	Q9CNY1	Q9cny1 pasteurella	635	6	4.3	307	10	Q9ASS3	Q9ass3 arabidopsis
563	6	4.3	286	16	Q8Z1E0	Q8z1e0 salmonella	636	6	4.3	308	16	Q8PGA3	Q8pga3 xanthomonas
564	6	4.3	286	16	Q8KCC8	Q8kcc8 chlorobium	637	6	4.3	308	17	Q9HPL3	Q9hpl3 halobacteri
565	6	4.3	287	16	Q9KTX0	Q9ktx0 vibrio chol	638	6	4.3	309	2	Q93UL0	Q93ul0 rhizobium g
566	6	4.3	287	16	Q8YPL6	Q8yp16 arabaeana sp	639	6	4.3	309	12	Q8V6M8	Q8v6m8 haeleivirus h
567	6	4.3	287	16	Q8UIM0	Q8uim0 agrobacteri	640	6	4.3	309	16	Q92SN5	Q92sn5 rhizobium m
568	6	4.3	287	17	P95897	P95897 sulfolobus	641	6	4.3	309	16	Q8DVI0	Q8dvr0 streptococ
569	6	4.3	288	10	Q9M1H2	Q9m1h2 arabidopsis	642	6	4.3	310	2	Q93UL1	Q93ul1 rhizobium g
570	6	4.3	288	16	Q9CHJ0	Q9chj0 lactococcus	643	6	4.3	310	10	Q8S4C4	Q8s4c4 zea mays m
571	6	4.3	288	16	Q99335	Q99335 mycobacteri	644	6	4.3	311	2	Q93UL6	Q93ul6 mesorhizobi
572	6	4.3	288	16	Q8DI77	Q8di77 synecococ	645	6	4.3	311	13	Q9DPF0	Q9ddf0 geochelone
573	6	4.3	289	10	Q95663	Q95663 closterium	646	6	4.3	311	10	Q8LRM2	Q8lrm2 petunia hyb
574	6	4.3	289	16	Q9A5Z9	Q9a5z9 caulobacter	647	6	4.3	312	11	Q9QUR4	Q9qur4 mus musculu
575	6	4.3	289	16	Q97DN3	Q97dn3 clostridium	648	6	4.3	312	11	Q9R2Z5	Q9r2z5 mus musculu
576	6	4.3	289	16	P96206	P96206 mycobacteri	649	6	4.3	312	16	Q8P4N5	Q8p4n5 xanthomonas
577	6	4.3	290	2	Q66013	Q66013 photobacter	650	6	4.3	313	2	Q8RNM4	Q8rnm4 legionella
578	6	4.3	290	4	Q9BSR0	Q9bsr0 homo sapien	651	6	4.3	313	2	Q86466	Q86466 xanthomonas
579	6	4.3	290	16	Q92SR4	Q92sr4 rhizobium m	652	6	4.3	313	11	Q8VFP0	Q8vfp0 mus musculu
580	6	4.3	290	16	Q92R86	Q92r86 vibrio vuln	653	6	4.3	313	16	Q82UY2	Q82uy2 rhizobium m
581	6	4.3	291	2	Q50170	Q50170 helicobacte	654	6	4.3	313	16	Q8UEA2	Q8uea2 agrobacteri
582	6	4.3	291	2	P94824	P94824 helicobacte	655	6	4.3	313	16	Q8PCA2	Q8pca2 xanthomonas
583	6	4.3	291	10	Q940W0	Q940w0 raphanus sa	656	6	4.3	313	17	Q8Z2Y2	Q8zzy2 pyrobaculum
584	6	4.3	291	16	Q9A2C8	Q9a2c8 caulobacter	657	6	4.3	314	16	Q92MU3	Q92mu3 rhizobium m
585	6	4.3	292	2	Q9K314	Q9k314 vibrio para	658	6	4.3	315	5	Q9N4G5	Q9n4g5 caenorhabdi
586	6	4.3	292	2	Q9K349	Q9k349 vibrio para	659	6	4.3	315	5	Q8T0Q2	Q8t0q2 thiosophila
587	6	4.3	292	2	Q9K309	Q9k309 vibrio para	660	6	4.3	315	16	Q884T5	Q884t5 rhizobium l
588	6	4.3	292	2	Q9K310	Q9k310 vibrio para	661	6	4.3	316	2	Q8KPR7	Q8kpr7 synecococ
589	6	4.3	292	4	Q8W566	Q8w566 homo sapien	662	6	4.3	316	16	Q9PHB4	Q9phb4 xylella fas
590	6	4.3	292	10	P93671	P93671 hordeum vul	663	6	4.3	316	16	Q8UHI2	Q8uhl2 agrobacteri
591	6	4.3	292	16	Q9ZKX6	Q9zkx6 helicobacte	664	6	4.3	316	16	Q8RA41	Q8ra41 thermotoga
592	6	4.3	292	16	Q9AK27	Q9ak27 streptomyc	665	6	4.3	317	16	Q9WTD4	Q9wdt4 thermotoga
593	6	4.3	292	16	Q8FRP4	Q8frp4 corynebacte	666	6	4.3	317	16	Q8DW20	Q8dw20 streptococ
594	6	4.3	292	16	Q8SUG8	Q8sug8 mycoplasma	667	6	4.3	318	10	Q9LOA5	Q9lqa5 arabidopsis
595	6	4.3	293	5	Q9MAU7	Q9mau7 drosophila	668	6	4.3	319	16	Q8XTB7	Q8xtb7 ralstonia s
596	6	4.3	293	10	Q23798	Q23798 triticum ae	669	6	4.3	320	16	Q8AF12	Q8af12 vibrio para
597	6	4.3	293	16	Q9PEB9	Q9peb9 xylella fas	670	6	4.3	321	2	Q93Z37	Q93z37 arabidopsis
598	6	4.3	294	5	Q8SQ14	Q8sq14 encephalito	671	6	4.3	322	5	Q45537	Q45537 caenorhabdi
599	6	4.3	294	5	Q8SRG2	Q8srg2 encephalito	672	6	4.3	322	16	Q8XBD0	Q8xbd0 escherichia
600	6	4.3	294	10	Q81987	Q81987 hordeum vul	673	6	4.3	322	16	Q8FH03	Q8fh03 escherichia

674	6	4.3	323	16	Q9A9T0	Q9a9t0 caulobacter	747	6	4.3	346	16	Q8RFZ3	Q8rfz3 fuobacteri
675	6	4.3	324	10	Q94WF7	Q94wf7 electrois sa	748	6	4.3	346	16	Q8ZDS8	Q8zds8 yersinia pe
676	6	4.3	324	8	Q8LM19	Q8lm19 oryza sativ	749	6	4.3	346	17	Q980J1	Q980j1 sulfolobus
677	6	4.3	325	15	Q9PD47	Q9pd47 xyella fas	750	6	4.3	347	3	Q04087	Q04087 saccharomyc
678	6	4.3	327	11	Q8C1K2	Q8clk2 mus musculu	751	6	4.3	347	4	Q16081	Q16081 homo sapien
679	6	4.3	327	16	Q98R62	Q98r62 mycoplasma	752	6	4.3	347	4	Q96NE4	Q96ne4 homo sapien
680	6	4.3	327	16	Q8RGU1	Q8rgu1 thermoanaer	753	6	4.3	347	5	Q9XVA5	Q9xva5 caenorhabdi
681	6	4.3	328	5	Q19421	Q19421 caenorhabdi	754	6	4.3	347	16	Q9CLB2	Q9clb2 pascuarella
682	6	4.3	329	2	Q9WX60	Q9wx60 acetobacter	755	6	4.3	347	16	Q8Z9M4	Q8z9m4 salmonella
683	6	4.3	329	16	Q92T16	Q92t16 rhizobium m	756	6	4.3	347	16	Q8Z9M4	Q8z9m4 salmonella
684	6	4.3	329	16	Q9RC56	Q9rc56 bacillus ha	757	6	4.3	348	2	Q9RHU2	Q9rhu2 streptomyc
685	6	4.3	329	16	Q8YBN4	Q8ybn4 streptococ	758	6	4.3	348	2	Q9RHU2	Q9rhu2 streptomyc
686	6	4.3	329	16	Q53927	Q53927 streptomyc	759	6	4.3	349	3	Q56775	Q56775 xanthomonas
687	6	4.3	330	10	Q80928	Q80928 arabidopsis	760	6	4.3	349	3	Q56775	Q56775 xanthomonas
688	6	4.3	331	2	Q8ZEU5	Q8zeu5 anabaena sp	761	6	4.3	349	16	Q8PNA2	Q8pna2 xanthomonas
689	6	4.3	331	2	Q52171	Q52171 pseudomonas	762	6	4.3	349	16	Q8PNA2	Q8pna2 xanthomonas
690	6	4.3	331	10	Q8S902	Q8s902 oryza sativ	763	6	4.3	349	16	Q8PNA2	Q8pna2 xanthomonas
691	6	4.3	331	13	Q73667	Q73667 gallus gall	764	6	4.3	350	16	Q98HK0	Q98hk0 rhizobium l
692	6	4.3	331	16	Q9X050	Q9x050 thermotoga	765	6	4.3	350	16	Q98HK0	Q98hk0 rhizobium l
693	6	4.3	331	16	Q8D2W7	Q8d2w7 wigglewort	766	6	4.3	350	16	Q8XWL2	Q8xwl2 ralstonia s
694	6	4.3	332	5	Q9NS15	Q9ns15 caenorhabdi	767	6	4.3	351	2	Q8LQ99	Q8lq99 stigmatella
695	6	4.3	332	10	Q9LE57	Q9le57 arabidopsis	768	6	4.3	351	3	Q9P959	Q9p959 emericella
696	6	4.3	332	11	Q9EQM7	Q9eqm7 mus musculu	769	6	4.3	351	17	Q8TQG7	Q8tgq7 methanosarc
697	6	4.3	332	12	Q914H9	Q914h9 sulfolobus	770	6	4.3	352	2	Q50851	Q50851 myxococcus
698	6	4.3	332	16	Q8NS28	Q8ns28 corynebacte	771	6	4.3	352	4	Q8NEE3	Q8nee3 homo sapien
699	6	4.3	332	16	Q9Z538	Q9z538 streptomyc	772	6	4.3	352	10	Q9LNV0	Q9lnv0 arabidopsis
700	6	4.3	333	2	Q9ZEK1	Q9zek1 anabaena va	773	6	4.3	352	10	Q8S614	Q8s614 oryza sativ
701	6	4.3	333	10	Q81494	Q81494 arabidopsis	774	6	4.3	352	16	Q8F4E4	Q8f4e4 xanthomonas
702	6	4.3	333	16	Q98C86	Q98c86 rhizobium l	775	6	4.3	352	17	Q8TKC6	Q8tkc6 methanosarc
703	6	4.3	333	16	Q8Z2X7	Q8z2x7 salmonella	776	6	4.3	353	5	Q9VNM1	Q9vnw1 drosophila
704	6	4.3	333	16	Q8YSU2	Q8ysu2 anabaena sp	777	6	4.3	353	10	Q8LDQ2	Q8ldq2 arabidopsis
705	6	4.3	333	16	Q8ZK02	Q8zk02 salmonella	778	6	4.3	353	10	Q8VZN6	Q8vzn6 arabidopsis
706	6	4.3	333	16	Q8D3B6	Q8d3b6 wigglewort	779	6	4.3	354	2	Q93L87	Q93l87 burkholderi
707	6	4.3	334	5	Q9TXK6	Q9txk6 caenorhabdi	780	6	4.3	354	3	Q8J123	Q8j123 emeticella
708	6	4.3	334	11	Q8R0U7	Q8rou7 mus musculu	781	6	4.3	355	2	Q31252	Q31252 anabaena va
709	6	4.3	334	16	Q8FRK9	Q8frk9 corynebacte	782	6	4.3	355	10	P93120	P93120 dianthua va
710	6	4.3	335	2	Q45728	Q45728 bacillus th	783	6	4.3	355	16	Q9XOV9	Q9xov9 thermotoga
711	6	4.3	335	10	Q9ZPM8	Q9zpm8 nepenthes a	784	6	4.3	356	3	Q9Y711	Q9y711 ajellomyces
712	6	4.3	335	10	Q8H954	Q8h954 nicotiana t	785	6	4.3	356	16	Q98PE0	Q98pe0 rhizobium l
713	6	4.3	335	16	Q8E541	Q8e541 streptococ	786	6	4.3	357	11	Q9DA61	Q9da61 mus musculu
714	6	4.3	335	16	Q8DZH5	Q8dzh5 streptococ	787	6	4.3	357	11	Q61784	Q61784 mus musculu
715	6	4.3	335	16	Q8PXU1	Q8pxu1 methanosarc	788	6	4.3	357	11	Q91V68	Q91v68 mus musculu
716	6	4.3	336	16	Q97TA9	Q97ta9 streptococ	789	6	4.3	357	11	Q60712	Q60712 mus musculu
717	6	4.3	337	5	Q8TOK8	Q8tok8 drosophila	790	6	4.3	358	3	Q12461	Q12461 saccharomyc
718	6	4.3	338	2	Q68474	Q68474 rhizobium s	791	6	4.3	358	3	Q96UR9	Q96ur9 monilia f
719	6	4.3	338	5	Q76655	Q76655 caenorhabdi	792	6	4.3	358	16	Q8XYX8	Q8xyx8 ralstonia s
720	6	4.3	338	16	Q9A611	Q9a611 caulobacter	793	6	4.3	359	3	P78762	P78762 schizosacch
721	6	4.3	338	16	Q98GG6	Q98gg6 rhizobium l	794	6	4.3	359	5	Q8IO70	Q8io70 drosophila
722	6	4.3	338	16	Q8ZJV1	Q8zjv1 salmonella	795	6	4.3	359	10	Q94J38	Q94j38 oryza sativ
723	6	4.3	339	11	Q9GYV4	Q9gyv4 rattus norv	796	6	4.3	359	12	Q84414	Q84414 paramedium
724	6	4.3	339	11	Q62708	Q62708 rattus norv	797	6	4.3	359	12	Q84414	Q84414 paramedium
725	6	4.3	339	16	Q8YE84	Q8ye84 brucella me	798	6	4.3	359	16	Q9KZ82	Q9kzh2 streptomyc
726	6	4.3	339	16	Q8P171	Q8p171 streptococ	799	6	4.3	359	16	Q8FRX8	Q8frx8 corynebacte
727	6	4.3	339	16	Q8K7M5	Q8k7m5 streptococ	800	6	4.3	361	2	Q8RKG8	Q8rkg8 clostridium
728	6	4.3	340	16	P72903	P72903 synecocyst	801	6	4.3	361	2	Q8RKG8	Q8rkg8 clostridium
729	6	4.3	340	16	Q8DRN6	Q8drn6 streptococ	802	6	4.3	361	2	Q8RKG4	Q8rkg4 clostridium
730	6	4.3	341	3	Q9Y7R8	Q9y7r8 schizosacch	803	6	4.3	361	2	Q8RKG2	Q8rk2 clostridium
731	6	4.3	341	10	Q9M3D5	Q9m3ds arabidopsis	804	6	4.3	361	2	Q8RKF8	Q8rkf8 clostridium
732	6	4.3	341	11	Q8VCN9	Q8vcn9 mus musculu	805	6	4.3	361	2	Q8RKF9	Q8rkf9 clostridium
733	6	4.3	341	16	Q92TS7	Q92ts7 rhizobium m	806	6	4.3	361	2	Q8RKG1	Q8rkg1 clostridium
734	6	4.3	341	16	Q9KYP7	Q9kyp7 streptomyc	807	6	4.3	361	2	Q8RKG6	Q8rkg6 clostridium
735	6	4.3	341	16	Q8DRZ4	Q8drz4 streptococ	808	6	4.3	361	2	Q8RKG7	Q8rk7 clostridium
736	6	4.3	342	2	Q52162	Q52162 pseudomonas	809	6	4.3	361	3	Q8NJ59	Q8nj59 botrytis ci
737	6	4.3	342	5	Q19282	Q19282 caenorhabdi	810	6	4.3	361	4	Q9UFQ1	Q9ufq1 homo sapien
738	6	4.3	343	10	Q9LQJ6	Q9lqj6 arabidopsis	811	6	4.3	362	2	Q99Q49	Q99q49 snigella fl
739	6	4.3	343	10	Q9XJ43	Q9xj43 atropa bell	812	6	4.3	362	10	Q9LJ34	Q9lj34 oryza sativ
740	6	4.3	344	2	Q8KZ07	Q8kz07 uncultured	813	6	4.3	362	12	Q91JES7	Q91jes7 crydia pomon
741	6	4.3	344	2	Q9L3U9	Q9l3u9 streptomyc	814	6	4.3	363	4	Q9H5G9	Q9h5g9 homo sapien
742	6	4.3	344	5	Q17990	Q17990 caenorhabdi	815	6	4.3	363	5	Q9GP32	Q9gp32 echinococcu
743	6	4.3	344	16	Q8NS84	Q8ns84 corynebacte	816	6	4.3	363	16	Q8FW54	Q8fm54 corynebacte
744	6	4.3	344	16	Q8FS07	Q8fs07 corynebacte	817	6	4.3	364	10	Q8RYR4	Q8ryr4 oryza sativ
745	6	4.3	345	10	Q8VXS9	Q8vxs9 beta vulgar	818	6	4.3	364	16	Q9CBQ4	Q9cbq4 mycobacteri
746	6	4.3	345	12	Q84439	Q84439 paramecium	819	6	4.3	364	16	P73423	P73423 synecocyst
													Q68573 streptococ

820	364	16	Q8F2A8	Q8f2a8 leptospira	893	6	4.3	392	13	Q9VHA5	Q9vha5 myxine glut
821	365	2	Q68776	Q68776 versinia pe	894	6	4.3	392	13	Q8AVB5	Q8avb5 cryzias lat
822	365	4	Q8NEW4	Q8new4 homo sapien	895	6	4.3	393	16	Q928F7	Q928f7 listeria in
823	365	16	Q9KfJ4	Q9kfj4 bacillus ha	896	6	4.3	393	16	O08395	O08395 mycobacteri
824	366	5	Q917Q7	Q917q7 drosophila	897	6	4.3	393	17	Q97YW4	Q97yw4 sulfolobus
825	366	11	Q8BX06	Q8bx06 mus musculu	898	6	4.3	394	2	Q9F241	Q9f241 xanthomonas
826	367	10	Q9SV98	Q9sv98 arabidopsis	899	6	4.3	394	2	Q9XBE6	Q9xbe6 amycolatops
827	368	10	Q9SB42	Q9sb42 arabidopsis	900	6	4.3	394	4	Q9Y3W7	Q9y3w7 homo sapien
828	369	2	Q9JRN1	Q9jrn1 actinobacil	901	6	4.3	394	6	O8H247	O8h247 felis silve
829	369	10	Q9SV45	Q9sv45 arabidopsis	902	6	4.3	394	16	P72780	P72780 synchocyst
830	369	16	Q9A3U6	Q9a3u6 caulobacter	903	6	4.3	395	2	Q9ZEG8	Q9zeg8 mycobacteri
831	369	16	Q8UGL7	Q8ugl7 agrobacteri	904	6	4.3	395	2	Q9RAJ8	Q9raj8 mycobacteri
832	370	5	Q44562	Q44562 caenorhabdi	905	6	4.3	395	2	Q9AKV5	Q9akv5 mycobacteri
833	370	5	Q94404	Q94404 caenorhabdi	906	6	4.3	395	11	Q9JHE0	Q9jhe0 rattus norv
834	370	16	Q89HP3	Q89hp3 rhizobium 1	907	6	4.3	395	17	Q97XP1	Q97xp1 sulfolobus
835	370	16	O8YKX9	O8ynk9 anabaena sp	908	6	4.3	395	17	O8TKY3	O8tky3 methanosarc
836	370	17	O29911	O29911 archaeoglob	909	6	4.3	395	17	O8TKY3	O8tky3 methanosarc
837	370	17	O27282	O27282 methanobact	910	6	4.3	396	2	O8LI60	O8li60 myxococcus
838	371	2	P86443	P86443 rhizobium m	911	6	4.3	396	4	Q96K21	Q96k21 homo sapien
839	371	2	O8GHQ5	O8ghq5 pseudomonas	912	6	4.3	396	4	O15564	O15564 trichomonas
840	371	5	Q9U2J4	Q9u2j4 caenorhabdi	913	6	4.3	396	5	Q9VVD9	Q9vvd9 drosophila
841	371	10	O81286	O81286 arabidopsis	914	6	4.3	396	5	O25553	O25553 naegleria f
842	371	11	Q9QWX9	Q9qwx9 mus musculu	915	6	4.3	396	10	Q94122	Q94122 oryza sativ
843	371	11	Q9QWY4	Q9qwY4 mus musculu	916	6	4.3	396	11	O61569	O61569 mus musculu
844	371	11	Q9QWX8	Q9qwx8 mus musculu	917	6	4.3	396	13	O8JHW6	O8jhw6 fugu rubrip
845	371	11	O54849	O54849 mus musculu	918	6	4.3	396	16	O8RI68	O8ri68 fuscobacteri
846	371	11	Q9QWY1	Q9qwY1 mus musculu	919	6	4.3	396	16	O8FM37	O8fm37 corynebacte
847	371	11	O88912	O88912 mus musculu	920	6	4.3	397	2	O9AGF0	O9agf0 mycobacteri
848	371	11	Q9QWY5	Q9qwY5 mus musculu	921	6	4.3	397	5	O8IEB4	O8ieb4 plasmodium
849	371	11	Q9QUJ6	Q9quj6 mus musculu	922	6	4.3	397	11	O8VDC4	O8vdc4 mus musculu
850	371	16	Q9PNS1	Q9pns1 campylobact	923	6	4.3	397	11	O9CWX7	O9cwx7 mus musculu
851	371	17	Q9HSF1	Q9hsf1 halobacteri	924	6	4.3	397	11	O8CFD5	O8cfd5 mus musculu
852	372	16	Q8PH06	Q8ph06 xanthomonas	925	6	4.3	397	16	Q9PH91	Q9ph91 xylella fas
853	372	16	Q8P5L2	Q8p5l2 xanthomonas	926	6	4.3	397	17	Q9V095	Q9v095 pyrococcus
854	374	10	Q9CSL2	Q9csl2 arabidopsis	927	6	4.3	398	3	Q9HED1	Q9hed1 neurospora
855	374	16	Q9X8Y8	Q9x8y8 streptomyce	928	6	4.3	398	11	O8CB15	O8cb15 mus musculu
856	374	17	Q97V53	Q97v53 sulfolobus	929	6	4.3	399	5	O8SQR3	O8sqR3 encephalito
857	374	17	Q8PZY2	Q8pzy2 methanosarc	930	6	4.3	399	9	O8W735	O8w735 bacterioph
858	376	4	O8WXH3	O8wxh3 homo sapien	931	6	4.3	399	10	O8LG40	O8lg40 arabidopsis
859	376	16	Q9S2T3	Q9s2t3 streptomyce	932	6	4.3	399	16	P74550	P74550 synchocyst
860	376	17	Q97XA4	Q97xa4 sulfolobus	933	6	4.3	399	17	Q97V47	Q97v47 sulfolobus
861	377	10	Q9FNC6	Q9fnc6 arabidopsis	934	6	4.3	400	15	O90KJ2	O90kj2 human immun
862	377	16	Q97EW8	Q97ew8 clostridium	935	6	4.3	401	16	Q97F24	Q97f24 clostridium
863	377	16	Q8XSG2	Q8xsg2 ralstonia s	936	6	4.3	401	17	O8PXS8	O8pxs8 methanosarc
864	378	10	Q9SVZ0	Q9svz0 arabidopsis	937	6	4.3	402	16	O8XST0	O8xst0 escherichia
865	378	16	O8EVB1	O8evb1 mycoplasma	938	6	4.3	402	16	O33312	O33312 mycobacteri
866	378	15	O8J4S1	O8j4s1 human immun	939	6	4.3	403	4	Q96CX6	Q96cx6 homo sapien
867	379	15	O8I760	O8i760 arabidopsis	940	6	4.3	403	5	Q9VCZ2	Q9vcz2 drosophila
868	381	4	O8NO22	O8no22 homo sapien	941	6	4.3	403	5	O8MZG6	O8mzg6 drosophila
869	381	5	Q9NE67	Q9ne67 caenorhabdi	942	6	4.3	403	10	O9FHS6	O9fhs6 arabidopsis
870	381	17	O58542	O58542 pyrococcus	943	6	4.3	403	17	Q97YI2	Q97yi2 sulfolobus
871	383	4	Q96DM5	Q96dm5 homo sapien	944	6	4.3	404	5	O20212	O20212 caenorhabdi
872	383	4	Q969Y6	Q969y6 homo sapien	945	6	4.3	404	10	Q941V4	Q941v4 oryza sativ
873	384	11	Q9EPB7	Q9epb7 mus musculu	946	6	4.3	404	12	Q98161	Q98161 little cher
874	384	11	O8B905	O8b905 mus musculu	947	6	4.3	404	12	O40962	O40962 little cher
875	384	11	Q9ESP4	Q9esp4 rattus norv	948	6	4.3	404	16	Q9EWI4	Q9ewi4 streptomyce
876	384	16	Q9A7D1	Q9a7d1 caulobacter	949	6	4.3	405	2	O8KT72	O8kt72 photorhabdu
877	385	5	O45442	O45442 caenorhabdi	950	6	4.3	406	2	O9RCB5	O9rcb5 versinia ps
878	385	5	O8IPB3	O8ipb3 drosophila	951	6	4.3	406	5	O9N4T4	O9n4t4 caenorhabdi
879	387	12	Q93887	Q93887 hepatitis b	952	6	4.3	406	10	Q9LYA9	Q9lya9 arabidopsis
880	387	16	Q9KY18	Q9kv18 vibrio chol	953	6	4.3	406	10	O8LA93	O8la93 arabidopsis
881	388	2	O8KHU2	O8khu2 actinosynne	954	6	4.3	406	10	O9LV70	O9lv70 arabidopsis
882	388	16	Q9K8K1	Q9k8k1 bacillus ha	955	6	4.3	407	2	O93AL9	O93al9 clostridium
883	388	16	O8FLY3	O8fly3 corynebacte	956	6	4.3	407	10	O22028	O22028 cyanidium c
884	389	2	Q54287	Q54287 streptomyce	957	6	4.3	407	10	O8GVH3	O8gvh3 arabidopsis
885	389	3	Q9HGL1	Q9hgl1 schizosacch	958	6	4.3	407	16	O8RJI9	O8rji9 streptomyce
886	390	2	Q9RH97	Q9rh97 salmonella	959	6	4.3	408	16	O9CJ33	O9cdj3 lactococcus
887	390	5	O9GVJ3	O9gvj3 caenorhabdi	960	6	4.3	408	17	Q97XQ8	Q97xq8 sulfolobus
888	390	10	Q9SR08	Q9sr08 arabidopsis	961	6	4.3	410	2	Q48368	Q48368 mycobacteri
889	390	16	O8XS34	O8xs34 ralstonia s	962	6	4.3	410	3	O12665	O12665 penicillium
890	391	10	P92929	P92929 antithamnio	963	6	4.3	411	2	O48882	O48882 mycobacteri
891	391	17	Q97Y44	Q97y44 sulfolobus	964	6	4.3	411	4	O8MW77	O8mw77 homo sapien
892	392	11	Q99KP5	Q99kp5 mus musculu	965	6	4.3	411	5	Q9VX69	Q9vx69 drosophila

Q9nf81 leishmania  
Q8sxg0 drosophila  
Q8u9u2 agrobacteri  
Q97v25 sulfolobus  
Q8ztr2 pyrobaculum  
Q9hfm0 metarhizium  
Q8fn18 corynebacte  
Q9rjg0 streptomyc  
Q8n836 homo sapien  
Q57456 synechocyst  
Q9z1t0 escherichia  
Q8kmw4 escherichia  
Q8wv18 homo sapien  
O15828 leishmania  
O62370 caenorhabdi  
O70629 mus musculu  
O8b691 mus musculu  
Q8fc97 escherichia  
Q9fsf2 nicotiana t  
P72207 pseudomonas  
Q9eg8 rhizobium l  
Q8emt7 oceanobacil  
Q8fvt2 brucella su  
Q94245 schizosacch  
Q9vnj9 drosophila  
Q93119 antheraea p  
Q9rn56 streptomyc  
Q9etg0 corynebacte  
Q9f826 micromonosp  
Q8i4h7 caenorhabdi  
Q68733 equine infe  
O67991 pseudomonas  
Q921s7 mus musculu  
Q91422 gallus gall  
Q8ah28 human immun

ALIGNMENTS

RESULT 1  
Q8MMN4  
ID Q8MMN4 PRELIMINARY; PRT; 141 AA.  
AC Q8MMN4; 141 AA.  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Putative vir15 protein.  
GN VIR.  
OS Babesia canis.  
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
OX NCBI\_TaxID=5867;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Drakulovski P., Carcy B., Moubri K., Carret C., Depoix D.,  
RA Schettters T.P.M., Gorenflot A.;  
RT "An extrachromosomal dRNA from Babesia canis implicated in parasite virulence."  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ494862; CAD41951.1; --  
SQ SEQUENCE 141 AA; 15752 MW; B84419C12BFD7CD1 CRC64;  
Query Match 100.0%; Score 141; DB 5; Length 141;  
Best Local Similarity 100.0%; Pred. No. 2.1e-136;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MESTSTTNFVAENRPTFGTFDVMREALLRVKSSERLALAGMCHRVLPOTGASA 60  
Db 1 MESTSTTNFVAENRPTFGTFDVMREALLRVKSSERLALAGMCHRVLPOTGASA 60  
Qy 61 IAATVTPKASMKLPPRPQSTKSPRLSKIRENMKTIQESARVNHRLPEGHPLLE 120  
Db 61 IAATVTPKASMKLPPRPQSTKSPRLSKIRENMKTIQESARVNHRLPEGHPLLE 120

Qy 121 KRAEYFRLRLSLKSGVNRLLI 141  
Db 121 KRAEYFRLRLSLKSGVNRLLI 141  
RESULT 2  
Q8MQH4  
ID Q8MQH4 PRELIMINARY; PRT; 285 AA.  
AC Q8MQH4;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Putative vir32 protein.  
GN VIR.  
OS Babesia canis.  
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
OX NCBI\_TaxID=5867;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Drakulovski P., Carcy B., Moubri K., Carret C., Depoix D.,  
RA Schettters T.P.M., Gorenflot A.;  
RT "An extrachromosomal dRNA from Babesia canis implicated in parasite virulence."  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ494862; CAD41950.1; --  
SQ SEQUENCE 285 AA; 32169 MW; 45ACEFEC7927243D CRC64;  
Query Match 84.4%; Score 119; DB 5; Length 285;  
Best Local Similarity 100.0%; Pred. No. 1.5e-113;  
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MESTSTTNFVAENRPTFGTFDVMREALLRVKSSERLALAGMCHRVLPOTGASA 60  
Db 1 MESTSTTNFVAENRPTFGTFDVMREALLRVKSSERLALAGMCHRVLPOTGASA 60  
Qy 61 IAATVTPKASMKLPPRPQSTKSPRLSKIRENMKTIQESARVNHRLPEGHPLL 119  
Db 61 IAATVTPKASMKLPPRPQSTKSPRLSKIRENMKTIQESARVNHRLPEGHPLL 119  
RESULT 3  
Q9LHSL  
ID Q9LHSL PRELIMINARY; PRT; 334 AA.  
AC Q9LHSL;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 22, Last annotation update)  
DE Genomic DNA, chromosome 5, TAC clone:K3D20.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids ii; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF002031; BAA98204.1; --  
DR InterPro; IPR002965; P-rich extensin.  
DR PRINTS; PR01217; PRICHEXTENS.  
SQ SEQUENCE 334 AA; 37849 MW; 5DF169A85F1B5FA9 CRC64;  
Query Match 5.7%; Score 8; DB 10; Length 334;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 59 SAIAATVT 66  
Db 20 SAIAATVT 27

Matches												8;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;	
RESULT 4																					
QY	Q9FG07	AC	Q9FG07	PRELIMINARY;	PRT;	434	AA.														
DB	14	SAIAATVT	21																		
Query Match												5.7%;	Score 8;	DB 10;	Length 434;						
Best Local Similarity												100.0%;	Pred. No. 17;								
Matches												8;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;	
QY	59	SAIAATVT	66																		
DB	14	SAIAATVT	21																		
RESULT 5																					
ID	Q9STM8	AC	Q9STM8	PRELIMINARY;	PRT;	513	AA.														
DT	01-MAY-2000	(TRENBLrel. 13, Created)																			
DT	01-MAY-2000	(TRENBLrel. 13, Last sequence update)																			
DT	01-MAR-2003	(TRENBLrel. 23, Last annotation update)																			
DE	Extensin-like protein.																				
GN	T28D5.90 OR AT4G08400.																				
OS	Arabidopsis thaliana (Mouse-ear cross).																				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;																				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;																				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.																				
OX	NCBI_TaxID=3702;																				
RN	[1]																				
RP	SEQUENCE FROM N.A.																				
RA	Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,																				
RA	Barrell B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K.,																				
RA	Schueeller C.;																				
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.																				
RN	[2]																				
RP	SEQUENCE FROM N.A.																				
RA	EU Arabidopsis sequencing project;																				
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.																				
RN	[3]																				
RP	SEQUENCE FROM N.A.																				
RA	Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,																				
RA	Mewes H.W., Lemcke K., Mayer K.F.X.;																				
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.																				
RN	[4]																				
RP	SEQUENCE FROM N.A.																				
RA	EU Arabidopsis sequencing project;																				
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.																				
DR	EMBL; AL109819; CAB52562.1;																				
DR	EMBL; AL161511; CAB37965.1;																				
SQ	SEQUENCE 513 AA; 57332 MW; D35888AA394E4401 CRC64;																				
Query Match												5.7%;	Score 8;	DB 10;	Length 513;						
Best Local Similarity												100.0%;	Pred. No. 19;								
Matches												8;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;	
Query Match												5.7%;	Score 8;	DB 16;	Length 534;						
Best Local Similarity												100.0%;	Pred. No. 20;								
Matches												8;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;	
QY	38	LAMLRALA	45																		
DB	151	LAMLRALA	158																		
RESULT 7																					
QY	Q9SK35	AC	Q9SK35	PRELIMINARY;	PRT;	559	AA.														
ID	Q9SK35	AC	Q9SK35	PRELIMINARY;	PRT;	559	AA.														
DT	01-MAY-2000	(TRENBLrel. 13, Created)																			
DT	01-MAY-2000	(TRENBLrel. 13, Last sequence update)																			
DT	01-OCT-2002	(TRENBLrel. 22, Last annotation update)																			
DE	At2G24980 protein.																				
GN	At2G24980.																				
OS	Arabidopsis thaliana (Mouse-ear cross).																				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;																				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;																				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.																				
OX	NCBI_TaxID=3702;																				
RN	[1]																				
RP	SEQUENCE FROM N.A.																				
RC	STRAIN=cv. Columbia;																				
RC	MEDLINE=20083497; PubMed=10617197;																				
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,																				
RA	Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,																				
RA	Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H., Moffat K.S.,																				
RA	Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,																				
RA	Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,																				
RA	Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,																				



RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana";  
 RL Nature 402:761-768 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Lin X.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AC06585; AAD23015.1; -  
 DR InterPro; IPR002965; P rich. extensn.  
 DR PRINTS; PR01217; PRICEXTENSN.  
 SQ SEQUENCE 559 AA; 63164 MW; B22AA763D0D1CC6D CRC64;

Query Match 5.7%; Score 8; DB 10; Length 559;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATVT 66  
 |||||  
 Db 14 SAIAATVT 21

RESULT 8  
 Q9FH26 PRELIMINARY; PRT; 609 AA.  
 ID Q9FH26;  
 AC Q9FH26;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Genomic DNA, chromosome 5, TAC clone:K20J1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=20181125; PubMed=10718197;  
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
 features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
 clones";  
 RL DNA Res. 7:31-63 (2000).  
 DR EMBL; AB023028; BAB10090.1; -  
 DR InterPro; IPR002965; P rich. extensn.  
 DR PRINTS; PR01217; PRICEXTENSN.  
 SQ SEQUENCE 609 AA; 68454 MW; C963750F7539DBA5 CRC64;

Query Match 5.7%; Score 8; DB 10; Length 609;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATVT 66  
 |||||  
 Db 14 SAIAATVT 21

RESULT 9  
 Q9FG06 PRELIMINARY; PRT; 689 AA.  
 ID Q9FG06;  
 AC Q9FG06;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Genomic DNA, chromosome 5, BAC clone:F15W7.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP002543; BAB11413.1; -  
 DR InterPro; IPR002965; P rich. extensn.  
 DR PRINTS; PR01217; PRICEXTENSN.  
 SQ SEQUENCE 689 AA; 77507 MW; 6AE991A17A2536C7 CRC64;

Query Match 5.7%; Score 8; DB 10; Length 689;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATVT 66  
 |||||  
 Db 39 SAIAATVT 46

RESULT 10  
 Q9STM7 PRELIMINARY; PRT; 707 AA.  
 ID Q9STM7;  
 AC Q9STM7;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Extensin-like protein.  
 GN T28D5.100 OR AT4G08410.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,  
 RA Barrrell B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K.,  
 RA Schueller C.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrrell B.G.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL109819; CAB52563.1; -  
 DR EMBL; AL161511; CAB77966.1; -  
 DR InterPro; IPR002965; P rich. extensn.  
 DR PRINTS; PR01217; PRICEXTENSN.  
 SQ SEQUENCE 707 AA; 79134 MW; E45D6B0E9FBFC032D CRC64;

Query Match 5.7%; Score 8; DB 10; Length 707;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SAIAATVT 66  
 |||||  
 Db 41 SAIAATVT 48

RESULT 11  
 Q9RDQ2 PRELIMINARY; PRT; 65 AA.  
 ID Q9RDQ2  
 AC Q9RDQ2;

DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DE Hypothetical protein SC02382.  
 GN SC02382 OR SC47.10.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=1902;  
 RN (1)\_TaxID=1902;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Harris D.E., Quail M.A., Kieser H.,  
 RA Thomson N.R., James K.D., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Harper D., Bateman A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Cronin A., Fraser A., Goble A., Larkel L., Murphy L., Oliver K., O'Neil S.,  
 RA Huang C.-H., Kieser T., Lake J., Larkel L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939112; CAB62714.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 65 AA; 6944 MW; F283FAL5A0650DCE CRC64;  
  
 Query Match 5.0%; Score 7; DB 16; Length 65;  
 Best Local Similarity 100.0%; Pred.No. 31;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 11 VAENRPT 17  
 DB 28 VAENRPT 34  
  
 RESULT 12  
 Q9BNJ8 PRELIMINARY; PRT; 91 AA.  
 ID Q9BNJ8  
 AC Q9BNJ8  
 DT 01-JUN-2001 (TREMELrel. 17, Created)  
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE RNA polymerase II largest subunit (Fragment).  
 OS Platydesmus sp. "pla".  
 OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Diplopoda; Helminthomorpha;  
 OC Platydesmida; Platydesmidae; Platydesmus.  
 OX NCBI\_TaxID=118492;  
 RN (1)\_TaxID=118492;  
 RP SEQUENCE FROM N.A.  
 RA Regier J.C., Shultz J.W.;  
 RT "A Phylogenetic analysis of Myriapoda (Arthropoda) using two nuclear  
 RT protein-encoding genes.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF240947; AAK11924.1; -;  
 DR InterPro; IPR007075; RNA\_pol\_Rpb1\_6.  
 DR Pfam; PF04992; RNA\_pol\_Rpb1\_6; 1.  
 FT NON\_TER 1  
 FT NON\_TER 91  
 SQ SEQUENCE 91 AA; 10270 MW; 7611D88426A06834 CRC64;  
  
 Query Match 5.0%; Score 7; DB 5; Length 91;  
 Best Local Similarity 100.0%; Pred.No. 43;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 88 RELSRKI 94  
 DB 22 RELSRKI 28  
  
 RESULT 13

Q49812 PRELIMINARY; PRT; 105 AA.  
 ID Q49812  
 AC Q49812  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
 DE B2168 C2 205.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Smith D.R.;  
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U00018; AAA17229.1; -;  
 SQ SEQUENCE 105 AA; 11450 MW; 24DDA109052F8DB0 CRC64;  
  
 Query Match 5.0%; Score 7; DB 2; Length 105;  
 Best Local Similarity 100.0%; Pred.No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 43 ALAGMCG 49  
 DB 42 ALAGMCG 48  
  
 RESULT 14  
 Q8DF92 PRELIMINARY; PRT; 133 AA.  
 ID Q8DF92  
 AC Q8DF92  
 DT 01-MAR-2003 (TREMELrel. 23, Created)  
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE Transcriptional regulator Crl-like protein.  
 OS Vibrio vulnificus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=672;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RT "Complete genome sequence of Vibrio vulnificus CMCP6.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE016798; AAC08856.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 133 AA; 15573 MW; 6B1538160953DB4F CRC64;  
  
 Query Match 5.0%; Score 7; DB 16; Length 133;  
 Best Local Similarity 100.0%; Pred.No. 60;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 83 KSPELRE 89  
 DB 51 KSPELRE 57  
  
 RESULT 15  
 Q49876 PRELIMINARY; PRT; 148 AA.  
 ID Q49876  
 AC Q49876  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE B229 F2 64.  
 OS Mycobacterium leprae.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1759;  
 [1]  
 RN RP SEQUENCE FROM N.A.  
 RA Smith D.R.;  
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U00020; AAA17318.1; -  
 SQ SEQUENCE 148 AA; 15679 MW; 9EC6126D28CE7E86 CRC64;  
 Query Match 5.0%; Score 7; DB 2; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 STSTTN 9  
 Db 85 STSTTN 91  
 RESULT 16  
 Q9VCT2 PRELIMINARY; PRT; 159 AA.  
 ID Q9VCT2  
 AC Q9VCT2;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE CG17244 protein (RH17470p).  
 GN CG17244.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen T.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.-H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003741; AAF56073.1; -  
 DR EMBL; AY113571; AAM29576.1; -  
 DR FlyBase; FBgn0039031; CG17244.  
 SQ SEQUENCE 159 AA; 17780 MW; 05EDC47AE278F315 CRC64;  
 Query Match 5.0%; Score 7; DB 5; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 ESTSTT 8  
 Db 29 ESTSTT 35  
 RESULT 17  
 Q9BDZ0 PRELIMINARY; PRT; 172 AA.  
 ID Q9BDZ0  
 AC Q9BDZ0;  
 DT 01-JUN-2001 (TRENBLrel. 17, Created)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Elastin (fragment).  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RA Sekhon H.S., Spindel E.R.;  
 RT "Maternal nicotine exposure up-regulates collagen and elastin gene  
 RT expression in fetal nonhuman primate lungs: potential role of alpha 7  
 RT nicotinic acetylcholine receptors.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF230927; AAK14974.1; -  
 DT NON\_TER 1  
 FT NON\_TER 172 172  
 SQ SEQUENCE 172 AA; 15224 MW; 4980EF4C5AF3330B CRC64;  
 Query Match 5.0%; Score 7; DB 6; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 52 VLPGTGA 58  
 Db 105 VLPGTGA 111  
 RESULT 18  
 Q9VH39 PRELIMINARY; PRT; 203 AA.  
 ID Q9VH39  
 AC Q9VH39;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE CG11722 protein (AT14909p).

GN CGL1722 OR BCDA:AT14909.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnikner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blake J.R., Champ M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celnik S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBDJ databases.  
 DR EMBL; AE003685; AAF54481.1; -;  
 DR EMBL; AY070793; AAL48415.1; -;  
 DR FlyBase; FBgn003777; CG11722.  
 SQ SEQUENCE 203 AA; 23705 MW; 325F385D98DC5F99 CRC64;  
 Query Match 5.0%; Score 7; DB 5; Length 203;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 29 LLRVKSS 35  
 Db 192 LLRVKSS 198  
 |||||  
 RESULT 19  
 Q95LB5

Q95LB5 PRELIMINARY; PRT; 223 AA.  
 AC Q95LB5;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Melanocortin-4 receptor (fragment).  
 GN MC4-R.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21421436; PubMed=11530231;  
 RA Iqbal J., Pompilio S., Dumont L.M., Wu C., Mountjoy K.G., Henry B.A.,  
 RA Clarke I.J.;  
 RT "Long-term alterations in body weight do not affect the expression of  
 RT melanocortin receptor-3 and -4 mRNA in the ovine hypothalamus."  
 RL Neuroscience 105:931-940(2001).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 DR EMBL; AF345913; AAL27186.1; -;  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm.1; 1.  
 DR PRINTS; PR00237; GPCRHOOPS.  
 DR PROSITE; PS00237; G-PROTEIN\_RECEP\_F1\_1; 1.  
 DR PROSITE; PS00262; G-PROTEIN\_RECEP\_F1\_2; 1.  
 DR G-protein coupled receptor; Receptor; Transmembrane.  
 FT NON TER 1  
 FT NON TER 223 223  
 SQ SEQUENCE 223 AA; 24542 MW; C8EBB5FBCAB766E CRC64;  
 Query Match 5.0%; Score 7; DB 6; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 52 VLPGTGA 58  
 Db 147 VLPGTGA 153  
 |||||  
 RESULT 20  
 Q8FWQ6  
 ID Q8FWQ6 PRELIMINARY; PRT; 225 AA.  
 AC Q8FWQ6;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Transcriptional regulator, GntR family.  
 GN BRA0389.  
 OS Brucella suis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29461;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1330 / Biovar 1;  
 RX MEDLINE=22247741; PubMed=12271122;  
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,  
 RA Read T.D., Dodson R.J., Umam L., Brinkac L.M., Beanan M.J.,  
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,  
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,  
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;  
 RT "The Brucella suis genome reveals fundamental similarities between  
 RT animal and plant pathogens and symbionts."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
 DR EMBL; AS014538; AAN3587.1; -;  
 DR TIGR; BRA0389; -;  
 SQ SEQUENCE 225 AA; 24928 MW; 4DA74F746593C4F0 CRC64;

```

Query Match          5.0%; Score 7; DB 16; Length 225;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 MREALLR 31
DB 57 MREALLR 63

RESULT 21
ID Q8YEM1 PRELIMINARY; PRT; 237 AA.
AC Q8YEM1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional regulator, GNTR family.
GN BMEII0878
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucellia.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756488;
RA Delveccio V.G., Kapatriel V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyripides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis";
RL PROC. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
DR EMBL; AB009721; AAL54120.1; -.
DR InterPro; IPR000524; HTH_GNTR.
DR Pfam; PF00392; gntr; 1.
DR SMART; SMO0345; HTH_GNTR; 1.
KW Complete proteome.
SQ SEQUENCE 237 AA; 26130 MW; 02655E0D6FE8BD65 CRC64;

Query Match          5.0%; Score 7; DB 16; Length 237;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 MREALLR 31
DB 69 MREALLR 75

RESULT 22
ID Q987C4 PRELIMINARY; PRT; 243 AA.
AC Q987C4;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional regulator.
GN MLR7110.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuono A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

```

```

RT Mesorhizobium loti ";
RL DNA Res. 7:331-338 (2000).
RL EMBL; AP003011; BAB53279.1; -.
DR InterPro; IPR000524; HTH_GNTR.
DR Pfam; PF00392; gntr; 1.
DR PRINTS; PR00035; HTHGNTR.
DR SMART; SMO0345; HTH_GNTR; 1.
DR PROSITE; PS00043; HTH_GNTR_FAMILY; 1.
KW Complete proteome.
SQ SEQUENCE 243 AA; 27217 MW; 72312C98217C4B22 CRC64;

Query Match          5.0%; Score 7; DB 16; Length 243;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 PPRPOST 82
DB 76 PPRPOST 82

RESULT 23
ID Q95SR4 PRELIMINARY; PRT; 255 AA.
AC Q95SR4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE GH05059p.
GN PPN OR CG1540 OR CG18436.
OS Drosophila melanogaster (Fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY060635; AAL28183.1; -.
DR FlyBase; FBgn0003137; Ppn.
SQ SEQUENCE 255 AA; 27727 MW; 99BF45863CFB1A63 CRC64;

Query Match          5.0%; Score 7; DB 5; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 ISOESAR 107
DB 172 ISOESAR 178

RESULT 24
ID Q93UL3 PRELIMINARY; PRT; 269 AA.
AC Q93UL3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nodulation protein C (fragment).
GN NODC.
OS Rhizobium giardinii bv. phaseoli.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=142630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H251;
RX PubMed=11283294;

```

RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
 RA Amarger N.;  
 RT "Classification of rhizobia based on nodC and nifH gene analysis  
 RT reveals a close phylogenetic relationship among Phaseolus vulgaris  
 symbionts.";  
 RL Microbiology 147:981-993(2001).  
 DR EMBL: AF217264; AAK39959.1; -.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR004835; Fungi\_chitin\_syn.  
 DR Pfam; PF03142; Chitin\_synth\_2; 1.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1 269  
 FT NON\_TER 269 269  
 SQ SEQUENCE 269 AA; 29825 MW; 02BBA4734884D175 CRC64;

Query Match 5.0%; Score 7; DB 2; Length 269;  
 Best Local Similarity 100.0%; Pred.No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 SAIATV 65  
 Db 148 SAIATV 154  
 |||||

RESULT 25  
 Q93UK8 PRELIMINARY; PRT; 270 AA.  
 AC Q93UK8  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Modulation protein C (fragment).  
 GN NODC.

OS Sinorhizobium sp. GR-06.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=147701;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GR-06;  
 RX PubMed=11283294;

RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
 RA Amarger N.;  
 RT "Classification of rhizobia based on nodC and nifH gene analysis  
 RT reveals a close phylogenetic relationship among Phaseolus vulgaris  
 symbionts.";  
 RL Microbiology 147:981-993(2001).  
 DR EMBL: AF217269; AAK39964.1; -.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR004835; Fungi\_chitin\_syn.  
 DR Pfam; PF03142; Chitin\_synth\_2; 1.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1 270  
 FT NON\_TER 270 270  
 SQ SEQUENCE 270 AA; 30089 MW; 5841B33D6EEDD131 CRC64;

Query Match 5.0%; Score 7; DB 2; Length 270;  
 Best Local Similarity 100.0%; Pred.No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 SAIATV 65  
 Db 154 SAIATV 160  
 |||||

RESULT 26  
 Q8RNL4 PRELIMINARY; PRT; 273 AA.  
 ID Q8RNL4  
 AC Q8RNL4

DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE NodC (fragment).  
 OS Sinorhizobium sp. 16b1.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=189051;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16b1;  
 RA Mhamdi R., Laguerre G., Aouani M.E., Mars M., Amarger N.;  
 RT "Different species and symbiotic genotypes of field rhizobia can  
 RT nodulate Phaseolus vulgaris in Tunisian soils.";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF481764; AAL88670.1; -.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR004835; Fungi\_chitin\_syn.  
 DR Pfam; PF03142; Chitin\_synth\_2; 1.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1 273  
 FT NON\_TER 273 273  
 SQ SEQUENCE 273 AA; 30557 MW; 2FFA2B2343E71325 CRC64;

Query Match 5.0%; Score 7; DB 2; Length 273;  
 Best Local Similarity 100.0%; Pred.No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 SAIATV 65  
 Db 160 SAIATV 166  
 |||||

RESULT 27  
 Q93UL5 PRELIMINARY; PRT; 281 AA.  
 AC Q93UL5  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Modulation protein C (fragment).  
 GN NODC.

OS Rhizobium etli bv. phaseoli.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=147700;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Viking1;  
 RX PubMed=11283294;

RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
 RA Amarger N.;  
 RT "Classification of rhizobia based on nodC and nifH gene analysis  
 RT reveals a close phylogenetic relationship among Phaseolus vulgaris  
 symbionts.";  
 RL Microbiology 147:981-993(2001).  
 DR EMBL: AF217262; AAK39957.1; -.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR004835; Fungi\_chitin\_syn.  
 DR Pfam; PF03142; Chitin\_synth\_2; 1.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1 281  
 FT NON\_TER 281 281  
 SQ SEQUENCE 281 AA; 31326 MW; D46A13EA185121C5 CRC64;

Query Match 5.0%; Score 7; DB 2; Length 281;  
 Best Local Similarity 100.0%; Pred.No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 SAIATV 65  
 Db 154 SAIATV 160  
 |||||

RESULT 28  
 Q8RNL4 PRELIMINARY; PRT; 273 AA.  
 ID Q8RNL4  
 AC Q8RNL4

```

SQ  SEQUENCE  283 AA;  30758 MW;  D2CEFI4B10653CFF  CRC64;

Query Match          5.0%;  Score 7;  DB 16;  Length 283;
Best Local Similarity 100.0%;  Pred. No. 1.2e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  27 EALLRVK 33
    |||||
DB  266 EALLRVK 272

RESULT 30
Q8U2B4
ID      Q8U2B4      PRELIMINARY;      PRT;      283 AA.
AC      Q8U2B4;
DT      01-JUN-2002 (TRENBLrel. 21, Created)
DT      01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT      01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DE      01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE      Heme biosynthesis protein.
DE      PF0925.
OS      Pyrococcus furiosus.
OS      Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC      Pyrococcus.
OC      NCBI_TaxID=2261;
RN      [1]_
RP      SEQUENCE FROM N.A.
RC      STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA      Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT      "The complete sequence of the Pyrococcus furiosus genome.";
RL      Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AS010206; AAL81049.1; -.
KW      Complete proteome.
SQ  SEQUENCE  283 AA;  32336 MW;  779EBE38D867B066  CRC64;

Query Match          5.0%;  Score 7;  DB 17;  Length 283;
Best Local Similarity 100.0%;  Pred. No. 1.2e+02;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

QY      83 ELSRKIR 95
      |||||
Db      273 ELSRKIR 279

RESULT 31
Q93UK9
ID      Q93UK9      PRELIMINARY;      PRG;      286 AA.
AC      Q93UK9;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DC      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Modulation protein C (Fragment).
OS      NODC.
OC      Rhizobium etli bv. phaseoli.
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC      Rhizobiaceae; Rhizobium.
CX      NCBI_TaxID=147700;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CfN42;
RX      PubMed=11283294;
RA      Laquerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,
RA      "Amarger N.,
RT      "Classification of rhizobia based on nodC and nifH gene analysis
RT      reveals a close phylogenetic relationship among Phaseolus vulgaris
RT      symbionts.";
RL      Microbiology 147:981-993(2001).
DR      EMBL; AF217268; AAK39963.1; -.
DR      InterPro; IPR001064; Crystallin.
DR      InterPro; IPR004935; Fungi_chitin_syn.
DR      InterPro; IPR001173; Glyco_transf_2.
DR      Pfam; PF03142; Chitin_synth_2; 1.
DR      Pfam; PF00535; Glycos_transf_2; 1.
DR      PROSITE; PS00325; CRYSTALLIN BETAGAMMA; 1.

```



FT NON\_TER 1 286  
FT NON\_TER 286  
SQ SEQUENCE 286 AA; 31784 MW; 90E168AB090F7B5B CRC64;  
Query Match 5.0%; Score 7; DB 2; Length 286;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 59 SAIAATV 65  
DB 161 SAIAATV 167  
RESULT 32  
Q93UL2 PRELIMINARY; PRT; 290 AA.  
AC Q93UL2;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Nodulation protein C (Fragment).  
GN NODC.  
OS Rhizobium gallium bv. phaseoli.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=142628;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Phd12;  
RX PubMed=11283294;  
RA Laquerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
Amarger N.;  
RT "Classification of rhizobia based on nodC and nifH gene analysis  
reveals a close phylogenetic relationship among Phaseolus vulgaris  
symbionts."  
RL Microbiology 147:981-993 (2001).  
DR EMBL; AF217265; AAK39960.1; -;  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR004835; Fungi\_chitin\_syn.  
DR InterPro; IPR001173; Glyco\_trans\_2.  
DR Pfam; PF03142; Chitin\_synth\_2; 1.  
DR Pfam; PF00535; Glycos\_transf\_2; 1.  
DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
FT NON\_TER 1 290  
FT NON\_TER 290  
SQ SEQUENCE 290 AA; 32213 MW; 7F9BD793B399F584 CRC64;  
Query Match 5.0%; Score 7; DB 2; Length 290;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 59 SAIAATV 65  
DB 165 SAIAATV 171  
RESULT 33  
O56780 PRELIMINARY; PRT; 297 AA.  
AC O56780;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Phosphoprotein.  
GN P.  
OS European bat lyssavirus 2.  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Rhabdoviridae; Lyssavirus.  
OX NCBI\_TaxID=57483;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=V286;  
RA Nadin-Davis S.A., Abdel-Malik M., Huang W., Armstrong J.,

Wandeler A.I.;  
RT "Genetic characterisation of the Lyssavirus P gene locus."  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF049121; AAC04591.1; -;  
DR InterPro; IPR004259; PP\_M1.  
DR Pfam; PF03012; PP\_M1; 1.  
SQ SEQUENCE 297 AA; 33287 MW; 52D56CSEA4902650 CRC64;  
Query Match 5.0%; Score 7; DB 12; Length 297;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 99 KTISOES 105  
DB 172 KTISOES 178  
RESULT 34  
Q9CIK6 PRELIMINARY; PRT; 308 AA.  
AC Q9CIK6;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Snz-type pyridoxine vitamin B6 biosynthetic protein SNZ1.  
GN PDX-1.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A;  
RX MEDLINE=21135672; PubMed=11238395;  
RA Bean L.E., Dvorachek W.H. Jr., Braun E.L., Errett A., Saenz G.S.,  
Giles M.D., Werner-Washburne M., Nelson M.A., Natvig D.O.;  
RT "Analysis of the pdx-1 (snz-1/sno-1) region of the Neurospora crassa  
genome. Correlation of pyridoxine-requiring phenotypes with mutations  
in two structural genes."  
RL Genetics 157:1067-1075 (2001).  
DR EMBL; AF309689; AAK07850.1; -;  
DR InterPro; IPR003009; FMN\_enzyme.  
DR InterPro; IPR001852; Snzlp/Sori.  
DR Pfam; PF01680; SOR\_SNZ; 1.  
DR ProDom; PD004958; Snzlp/Sori; 1.  
DR TIGRFAMs; TIGR00343; TIGR00343; 1.  
DR PROSITE; PS01235; UPF0019; 1.  
SQ SEQUENCE 308 AA; 32355 MW; 47D73BB4CAD284D7 CRC64;  
Query Match 5.0%; Score 7; DB 3; Length 308;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 STSTTN 9  
DB 3 STSTTN 9  
RESULT 35  
Q93UK6 PRELIMINARY; PRT; 311 AA.  
AC Q93UK6;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Nodulation protein C (Fragment).  
GN NODC.  
OS Rhizobium leguminosarum (biovar trifolii).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=386;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=USDA2071;  
 RX PubMed=11283294;  
 RA Laguerre G., Nour S.M., Macheret V., Sanjuan J., Drouin P.,  
 RA Amarger N.,  
 RA "Classification of rhizobia based on nodC and nifH gene analysis.  
 RT reveals a close phylogenetic relationship among Phaseolus vulgaris  
 RT symbionts.";  
 RL Microbiology 147:981-993(2001).  
 DR EMBL: AF217271; AAK39966.1; -.  
 DR InterPro: IPR001064; Crystallin.  
 DR InterPro: IPR004835; Fungi chitin syn.  
 DR InterPro: IPR001173; Glyco-trans\_2.  
 DR Pfam: PF03142; Chitin synthase 2; 1.  
 DR Pfam: PF00335; Glycosyl transferase 2; 1.  
 DR PROSITE: PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1 1  
 FT 311 311  
 SQ SEQUENCE 311 AA; 34360 MW; B3C03D399D59081B CRC64;  
 Query Match 5.0%; Score 7; DB 2; Length 311;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 SAIAATV 65  
 DB 176 SAIAATV 182  
 RESULT 36  
 Q8XGX9 PRELIMINARY; PRT; 323 AA.  
 ID Q8XGX9  
 AC Q8XGX9  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Preprotein translocase, IISF family, membrane subunit (Protein-export  
 DE membrane protein SecY).  
 GN SECY OR STM0408 OR STY0446.  
 OS Salmonella typhimurium, and  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602, 601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking I., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2.";  
 RL Nature 413:852-855(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhi; STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrall B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 DR EMBL: AE008714; AAL19362.1; -.  
 DR EMBL: AL627266; CAD08864.1; -.  
 DR InterPro: IPR003395; SecD\_SecF.

DR InterPro: IPR005665; SecF.  
 DR Pfam: PF02355; SecD\_SecF; 1.  
 DR TIGRFAMs: TIGR00916; 2A0604801; 1.  
 DR TIGRFAMs: TIGR00966; 3A0S01807; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 323 AA; 35376 MW; 10D06394DF60CB86 CRC64;  
 Query Match 5.0%; Score 7; DB 16; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 DVMREAL 29  
 DB 69 DVMREAL 75  
 RESULT 37  
 Q9I4V9 PRELIMINARY; PRT; 330 AA.  
 ID Q9I4V9  
 AC Q9I4V9  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Probable glycosyl transferase.  
 GN PA1014.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
 RA Hickey M.J., Brinkman F.S.L., Huynh W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzer S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Brody L.L., Coulter S.N., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004534; AAG04403.1; -.  
 DR HSP; P39621; IQGQ.  
 DR InterPro: IPR01173; Glyco-trans\_2.  
 DR Pfam: PF00535; Glycosyl transferase 2; 1.  
 KW Transferase; Complete proteome.  
 SQ SEQUENCE 330 AA; 37555 MW; 77F8654AE29F86 CRC64;  
 Query Match 5.0%; Score 7; DB 16; Length 330;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 36 ERLAMLR 42  
 DB 79 ERLAMLR 85  
 RESULT 38  
 Q9AD29 PRELIMINARY; PRT; 331 AA.  
 ID Q9AD29  
 AC Q9AD29  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein SCPI121.  
 GN SCPI.121.  
 OS Streptomyces coelicolor.  
 OG Plasmid SCPI.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]

```

Query Match      5.0%; Score 7; DB 6; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches      7; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      52 VLPGTGA 58
      |||||
Db      228 VLPGTGA 234

RESULT 40
Q97YG7      PRELIMINARY;      PRT;      332 AA.
ID      Q97YG7;
DT      01-OCT-2001 (TrEMBLrel. 18, Created)
DT      01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT      01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE      Endoglucanase precursor (EC 3.2.1.4).
GN      SSO1354.
OS      Sulfolobus solfataricus.
OC      Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC      Sulfolobus.
OX      NCBI_TaxID=2287;
[1]
RN      SEQUENCE FROM N.A.
RP
RC      STRAIN=ATCC 35092 / DSM 1617 / P2;
RX      MEDLINE=21332296; PubMed=11427726;
RA      She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA      Awatey M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA      De Moors A., Brauso G., Fletcher C., Gordon P.J.K.,
RA      Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA      Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA      Charlebois R.B., Doolittle W.F., Duguet M., Gaasterland T.,
RA      Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT      "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RT      Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RL      EMBL; AE006749; AAK41590.1;
DR      InterPro; IPR002594; Glyco_hydro_12.
DR      Pfam; PF01670; Glyco_hydro_12; 1.
DR      ProDom; PD004316; Glyco_hydro_12; 1.
DR      Hydrolase; Glycosidase; Complete proteome.
SQ      SEQUENCE      332 AA;      37295 MW;      F896165AE4926B9E      CRC64;

Query Match      5.0%; Score 7; DB 17; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches      7; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      3 STSTTTN 9
      |||||
Db      45 STSTTTN 51

Search completed: November 14, 2003, 10:52:09
Job time : 65 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 17, 2003, 12:05:08 ; Search time 2089 seconds  
(without alignments)  
2761.253 Million cell updates/sec

Title: US-10-087-573-2

Perfect score: 708

Sequence: 1 MESTSTTTNFVAENRPTFGE.....RAEYFRHLRLSLKSGQGNRLI 141

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh

-Q/cgn2\_1/USPTO.spool/US10087573/runat\_14112003\_103610\_20673/app.query.fasta\_1.327

-DB=GenEmbl -OPT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45

-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL

-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US10087573 @cgn 1 1 2372 @runat\_14112003\_103610\_20673 -RCPU=3

-NO WAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-FV TIMEOUT=120 -WARN. TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6

-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: 1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	708	100.0	1115	3	BCA494862
2	708	100.0	1135	6	AX530469
3	680	96.0	1134	6	AX530471
4	95.5	13.5	10463	1	AE012139
5	92.5	13.1	4656	9	HSM803142
6	92.5	13.1	4839	6	AX430339
7	92.5	13.1	5973	1	AB024056
8	92.5	13.1	6573	1	AB027763
9	92.5	13.1	10123	6	AX367101
10	92.5	13.1	10759	6	BD084093
11	92.5	13.1	10759	9	AB053446
12	92.5	13.1	192239	9	AC009796
13	92.5	13.1	208430	2	AC027641
14	92.5	13.1	211735	9	AC091564
15	92	13.0	90	6	AX530473
16	92	13.0	11060	1	AE004618
17	91.5	12.9	1386	10	AF416730
18	91.5	12.9	1607	10	BC002136
19	91.5	12.9	1618	10	AF495469
20	91.5	12.9	1633	10	AF143955
21	90	12.7	1563	9	HSCORONIN
22	90	12.7	1589	9	HSU34690
23	90	12.7	1639	9	AF495470
24	90	12.7	1657	6	E11332
25	90	12.7	1659	6	AX237691
26	90	12.7	1659	9	HUMP57B
27	90	12.7	208443	2	AC120207
28	90	12.7	210614	1	AB088224
29	90	12.7	217346	9	AC012363
30	89.5	12.6	1227	1	TFEMADF
31	89	12.6	96236	3	AC012492
32	88.5	12.5	176562	9	AC084864
33	88.5	12.5	198677	1	AE001863
34	88	12.4	846	6	BD180073
35	88	12.4	148797	2	AP005426
36	88	12.4	156242	2	AP005429
37	88	12.4	162382	2	AP005180
38	88	12.4	283100	1	SC0939110
39	88	12.4	300813	2	AC125900
40	88	12.4	308821	2	AC118450
41	88	12.4	348430	2	AC130157
42	87.5	12.4	1386	10	AF047388
43	87.5	12.4	1605	10	BC053398
44	87.5	12.4	1616	10	AF495468
45	87.5	12.4	2115	5	BC047956

# ALIGNMENTS

RESULT 1

BCA494862 1115 bp mRNA linear INV 25-FEB-2003  
LOCUS Babesia canis vir gene for putative vir15 and vir32 proteins,  
DEFINITION genomic RNA.  
ACCESSION AJ494862  
VERSION AJ494862.1 GI:21748167  
KEYWORDS ribosomal slippage; vir gene; vir15 protein; vir32 protein.  
SOURCE Babesia canis  
ORGANISM Babesia canis  
Eukaryota; Alveolata; Apicomplexa; Pirolasmida; Babesiidae;  
Babesia.  
REFERENCE 1  
AUTHORS Drakulovski P., Carcy B., Moubri K., Carret C., Depoix D.,  
Schettters T.P. and Gorenflot A.  
TITLE Antibodies raised against BcVir15, an extrachromosomal  
double-stranded RNA-encoded protein from Babesia canis, inhibit the  
in vitro growth of the parasite  
JOURNAL Infect. Immun. 71 (3), 1056-1067 (2003)  
MEDLINE 22483655  
PUBMED 12595415  
REFERENCE 2 (bases 1 to 1115)  
AUTHORS Drakulovski P.  
TITLE Direct Submission  
JOURNAL Submitted (10-JUL-2002) Drakulovski P., Lab de Biologie Cellulaire  
et Molculaire, UFR des Sciences Pharmaceutiques et Biol, 15 av  
Charles Flahaut, BP 14491, F-34093 Montpellier Cedex 5, FRANCE  
FEATURES  
source  
1. .1115  
/organism="Babesia canis"  
/mol\_type="mRNA"  
/db\_xref="taxon:5867"  
/notes="extrachromosomal dsRNA"  
61. .919  
/genes="vir"  
join(61..417,419..919)  
/genes="vir"  
/function="putative function in virulence"  
/notes="ribosomal slippage"  
/codon\_start=1  
/evidence=experimental  
/product="putative vir32 protein"  
/protein\_id="CAD41950.1"  
/db\_xref="GI:21748168"  
/db\_xref="SPTREMBL:Q8VMH4"  
/translation="MSTSTTTNFVAENRPTFGTDFVMEALLRVKSSERLAMLRL  
AGMCHRVLPCTGASAIATVTPKGSMLKPPRQSTKSPRLRLSKIRKMNKTIS  
QESARVNHLPPEGHPLRSGQNIPTVILDLRAKESIDSKALRRYFRASMRNTRLVH  
NRPVLPKVEPDSNLPFGQRSMRTWNRPRTPGYSNGTLAVTELLISIVRSNFT  
LKVEGRCTCCNTHKEQALLSLGYLDYLRALHVSGRSVFVEYCKTRICVEARLTGL  
RPRTVLTGC"  
61. .486  
/genes="vir"  
/function="putative function in virulence"  
/codon\_start=1  
/evidence=experimental  
/product="putative vir15 protein"  
/protein\_id="CAD41951.1"  
/db\_xref="GI:21748169"  
/db\_xref="SPTREMBL:Q8VMN4"  
/translation="MSTSTTTNFVAENRPTFGTDFVMEALLRVKSSERLAMLRL  
AGMCHRVLPCTGASAIATVTPKGSMLKPPRQSTKSPRLRLSKIRKMNKTIS  
QESARVNHLPPEGHPLRSGQNIPTVILDLRAKESIDSKALRRYFRASMRNTRLVH  
NRPVLPKVEPDSNLPFGQRSMRTWNRPRTPGYSNGTLAVTELLISIVRSNFT  
LKVEGRCTCCNTHKEQALLSLGYLDYLRALHVSGRSVFVEYCKTRICVEARLTGL  
RPRTVLTGC"  
BASE COUNT 303 a 281 c 282 g 249 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6.93e-53 Length: 1115  
Score: 708.00 Matches: 141  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-10-087-573-2 (1-141) x BCA494862 (1-1115)  
QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 61 ATGGAGTCGACATCAACCAACGACCACTTTGTCGCGAGAACCGTCCACCTTTGGTCAG 120  
QY 21 ThrPheAspValMetArgGluAlaLeuLeuArgVallySerSerGluArgLeuAlaMet 40  
Db 121 ACGTTTGATGTGATGAGGGAAGCTTTGCTTGTGTAAAGTCTCTGAAACGCTTGCAATG 180  
QY 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyValAsnAla 60  
Db 181 CTCAGAGCGCTTGCAGGAATGTGCGGTCAACCGCGTCTCTCGCACCTGGTCTCTCGG 240  
QY 61 IleAlaAlaThrValThrProlyGlyAlaSerMetIlySerLeuLysProProArgProGln 80  
Db 241 ATAGCGGCAACGGTAACCCCAAGGGGCTTCGATGAAGCTTAAACCAACCGCTCCGCGAG 300  
QY 81 SerThrLysSerProGluLeuArgGluLeuSerArgIlyIleArgGluMetAsnLysThr 100  
Db 301 TCAACGAAGTCTCCGAGCTCAGGAGCTGTACGGAAGATTCGCGAATGAATAAGACT 360  
QY 101 IleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGlu 120  
Db 361 ATAAATCAGGAATCAGCTCGGTAAACCAACCGGTTCGCGAAGCCACCTCTCTTAGAG 420  
QY 121 LysArgAlaGluTyThrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 421 AAGCGGCGAGAAATATTTTCGTACCTTAGATCTCTTAAGACCAAGGAGTCAATAGACTC 480  
QY 141 Ile 141  
Db 481 ATC 483  
RESULT 2  
AX530469 1135 bp DNA linear PAT 22-NOV-2002  
LOCUS Sequence 1 from Patent EP1238983.  
DEFINITION AX530469  
ACCESSION AX530469  
VERSION AX530469.1 GI:25252327  
KEYWORDS Babesia canis  
SOURCE Babesia canis  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Pirolasmida; Babesiidae;  
Babesia.  
REFERENCE 1  
AUTHORS Schettters T.P., Carcy B.P., Drakulovski P.R. and Gorenflot A.F.  
TITLE Babesia canis vaccine  
JOURNAL Patent: EP 1238983-A 1 11-SEP-2002;  
Akzo Nobel N.V. (NL)  
FEATURES  
source  
1. .1135  
/organism="Babesia canis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:5867"  
75. .500  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD57205.1"  
/db\_xref="GI:25252328"  
/translation="MESTSTTTNFVAENRPTFGTDFVMEALLRVKSSERLAMLRL  
AGMCHRVLPCTGASAIATVTPKGSMLKPPRQSTKSPRLRLSKIRKMNKTIS  
QESARVNHLPPEGHPLRSGQNIPTVILDLRAKESIDSKALRRYFRASMRNTRLVH  
NRPVLPKVEPDSNLPFGQRSMRTWNRPRTPGYSNGTLAVTELLISIVRSNFT  
LKVEGRCTCCNTHKEQALLSLGYLDYLRALHVSGRSVFVEYCKTRICVEARLTGL  
RPRTVLTGC"  
BASE COUNT 308 a 286 c 289 g 252 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.08e-53 Length: 1135  
Score: 708.00 Matches: 141  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

```

US-10-087-573-2 (1-141) x AX530469 (1-1135)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20
   |||||
Db 75 ATGGAGTCGACATCAACACGACCACTTTGTTGCCGAGAACCGTCCACCTTTGGTGAG 134

QY 21 ThrPheAspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMet 40
   |||||
Db 135 ACCTTTTGATGTGATGAGGAGGAGCTTTGCTTCGTGTAAGTCTCTCTGAACGCTTGCGCAATG 194

QY 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60
   |||||
Db 195 CTCAGAGCGCTTCGAGGAATGTCCGGTCAACCGGCTCTTCTGCGACTGGTCTTCTGCG 254

QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80
   |||||
Db 255 ATAGCGGCAACGGTAAACCCCAAGGGGGCTTCGATGAAGCTTAAACACCGCGTCCGCGAG 314

QY 81 SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr 100
   |||||
Db 315 TCAACGAAGTCTCCGAGCTCAGGAGCTGTCAACGAGATTCGCGAATGAATAGACT 374

QY 101 IleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGlu 120
   |||||
Db 375 ATAAGTCAGGAATCAGCTCGGTTAAACACCGGTTGCCGAGGACCCCTCTCTTAGAG 434

QY 121 LysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140
   |||||
Db 435 AAGCGGGCAGAAATATTTCGTCACTTAGATCTCTTAGACCAAGGAGTCAATAGACTC 494

141 Ile 141
   |||
495 ATC 497

RESULT 3
AX530471 1134 bp DNA linear PAT 22-NOV-2002
LOCUS
DEFINITION
Sequence 3 from Patent EP1238983.
ACCESSION
AX530471
VERSION
AX530471.1 GI:25252330
KEYWORDS
Babesia canis
SOURCE
Babesia canis
ORGANISM
Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae;
Babesia.

REFERENCE
1 Schetters,T.P., Carcy,B.P., Drakulovski,P.R. and Gorenflot,A.F.
  Babesia canis vaccine
  Patent: EP 1238983-A 3 11-SEP-2002;
  Akzo Nobel N.V. (NL)

FEATURES
Location/Qualifiers
1..1134
   /organism="Babesia canis"
   /mol_type="genomic DNA"
   /db_xref="taxon:5867"
75..932
   /note="unnamed protein product"
   /codon_start=1
   /protein_id="CAD57206.1"
   /db_xref="GI:25252331"
   /translation="MESTSTTNFVAENRPTFGTDFVMEALLRVKSSERLAMLRAL
AGMCHRLVPTGASAIATVTPKASMKLPPRQSTKSPKPELRSKIRKEMNKTIS
CESAVNRHLPGEHPLLEKRAEYFTVLDLRAKESIDSKLRRYRASVRNTRLVH
NRRVLKVPDSDNLPFGQRRSRMTWNLRPRGTGYSNGTLAVTELLISVSNFT
LKVVEGRGTCNCTHKQALLLSGLYLQLYALHLSVGRSVFVYCKTRICVEARLGL
RPRVTLGTC"

BASE COUNT 308 a 286 c 289 g 251 t
ORIGIN

Alignment Scores:
Pred. No.: 1134
Score: 680.00 Matches: 140

```

```

Percent Similarity: 99.29% Conservative: 0
Best Local Similarity: 99.29% Mismatches: 1
Query Match: 96.05% Indels: 1
DB: 6 Gaps: 0

US-10-087-573-2 (1-141) x AX530471 (1-1134)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20
   |||||
Db 75 ATGGAGTCGACATCAACACGACCACTTTGTTGCCGAGAACCGTCCACCTTTGGTGAG 134

QY 21 ThrPheAspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMet 40
   |||||
Db 135 ACCTTTTGATGTGATGAGGAGGAGCTTTGCTTCGTGTAAGTCTCTCTGAACGCTTGCGCAATG 194

QY 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60
   |||||
Db 195 CTCAGAGCGCTTCGAGGAATGTCCGGTCAACCGGCTCTTCTGCGACTGGTCTTCTGCG 254

QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80
   |||||
Db 255 ATAGCGGCAACGGTAAACCCCAAGGGGGCTTCGATGAAGCTTAAACACCGCGTCCGCGAG 314

QY 81 SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr 100
   |||||
Db 315 TCAACGAAGTCTCCGAGCTCAGGAGCTGTCAACGAGATTCGCGAATGAATAGACT 374

QY 101 IleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGlu 120
   |||||
Db 375 ATAAGTCAGGAATCAGCTCGGTTAAACACCGGTTGCCGAGGACCCCTCTCTTAGAG 434

QY 121 LysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140
   |||||
Db 435 AAGCGGGCAGAAATATTTCGTCACTTAGATCTCTTAGACCAAGGAGTCAATAGACTC 493

141 Ile 141
   |||
494 ATC 496

RESULT 4
AX530471 10463 bp DNA linear BCT 23-MAY-2002
LOCUS
DEFINITION
Xanthomonas campestris pv. campestris str. ATCC 33913, section 47
of 460 of the complete genome.
ACCESSION
AX530471
VERSION
AX530471.1 GI:21111398
KEYWORDS
Xanthomonas campestris pv. campestris str. ATCC 33913
Xanthomonas campestris pv. campestris str. ATCC 33913
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
1 (bases 1 to 10463)
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C. do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S.,
Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
Meneses,C.F.N., Miyaki,C.I., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,K.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
Rossa,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinoia,L.A.F.,
Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kitajima,J.P.
Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities
Nature 417 (6887), 459-463 (2002)
JOURNAL
22022145
MEDLINE
12024217
PUBMED

```

REFERENCE	2 (bases 1 to 10463)		
AUTHORS	da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,W.C., Camargo,L.B.A., Camarotte,G., Cannavan,F., Cardozo,J., Chamergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.B., Cursino-Santos,J.R., El-Dorri,H., Faria,U.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezra,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-NOV-2001) Departamento de Bioquimica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil		
FEATURES	Location/Qualifiers		
source	1..10463		
	/organism="Xanthomonas campestris pv. campestris str. ATCC 33913"		
	/mol_type="genomic DNA"		
	/strain="ATCC 33913"		
	/db_xref="ATCC:33913"		
	/db_xref="taxon:190485"		
	/note="pathovar: campestris"		
	/complement(457..1215)		
	/genes="XCC0416"		
	/complement(457..1215)		
	/genes="XCC0416"		
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"		
	/codon_start=1		
	/transl_table=11		
	/product="3-oxoacyl-[ACP] reductase"		
	/protein_id="AAM39734.1"		
	/db_xref="GI:21111399"		
	/translation="WSTPPQPVRLVIGSGIGLATAQFAVQSGAQSVCARNPEG LANAQAALDAGPVHTFACDLADAQERYVHAQAQFGLDVVNVNAGSHGND ESWQGLVDLMAVRCNAVPYLRQSGNAVILNISSINGRPTFPRIAYSTAKAAL NYITTLAELAREIRVNAIPAGSIEFEGWLEQRSDEPALYARIDSIPFGFGQ VQIADAALFLASPOARWITGQVLAVDQGSILGV"		
gene	1518..2954		
	/gene="XCC0417"		
	1518..2954		
	/gene="XCC0417"		
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"		
	/codon_start=1		
	/transl_table=11		
	/product="conserved hypothetical protein"		
	/protein_id="AAM39735.1"		
	/db_xref="GI:21111400"		
	/translation="VQVFERLHMETHTVIGAGCGSVLVRALAQGADAQVRLTVG VAEFCSGIAYGAARPEHLNVRKLDGIDAQPAAFADTLHGESRLEFLPRIAY DYRLSELDITGAQAQASIVRLSQEAVAVERRGRFVFLANGDAFQTDQVVLVAGLQ PQALAGIPRLGPHRYIGVPWQDALSPPDDVLIVGTLTMDVALSLHAGHR GRLLATSRGLAPQALHQPAALELPHLQRAIKDADRLALRGVROLVVVDWR VVDALRPHLQPLWQLELSORAFRLHLRPYEVARHVPAGADQLAQDLSQLOV VAARLLRAVPDGVGEAVIRPFGGADQTRFDVAVRATGLDITDRTSDLLAGMRE AGLLRADPLGLVDITDAOLRVDPDQGVQVGLYCVGPLLRGRYWEITAVPELRVATRA LAERLLQATPALQFQPDVTPREVNAP"		
gene	3277..3708		
	/gene="virk"		
	/note="synonym: XCC0418"		
	3277..3708		
	/gene="virk"		
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"		
	/codon_start=1		
	/transl_table=11		
	/product="cation efflux system protein"		
	/protein_id="AAM39736.1"		
	/db_xref="GI:21111401"		
	/translation="MFRPASIVTAALCACALPALASETNSRTDVLAALEAGYDVSV STDLRSTPPEGTPVSTGRGHIDAVRITEDGNVAFSGAHFTVANDGKPTQOPMRVQ LRPDGSVRFTTYMDLQGLQGRGPVLAYQCAFNEGAFHAD"		
gene	complement(3903..5375)		
	/gene="oprN"		
	/note="synonym: XCC0419"		
	complement(3903..5375)		
	/gene="oprN"		
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"		
	/codon_start=1		
	/transl_table=11		
	/product="outer membrane protein"		
	/protein_id="AAM39737.1"		
	/db_xref="GI:21111402"		
	/translation="MPSRLTLVLLTASLASCAPVEYARPDAPLPORYLAQDAEQR SAAPVDLSRWAGFGDPALDRVYVTAALQNLDAQAAAVTOARGLGAANAALLPS LINGQATRYQSLETPDGOVLNTPDFDRYASVYDATANASWEIDVFGILRGROA LADYQASAAGAATRLAVAQAQADTVISIRGLQARLAVAGROVATQOQLAMVTLLOG KGLAADLVQQAQALVQAQGVPALETALTAKMNLDMVGATPGTHRAELVASTPI PTAQVTAASPGDLLRRRDLVAERRLAVANARIGVMAEYYPKFSLSGLIGSATA ISGANLFTGDAQASAGVLRLWLFDFGRINAQIAQAKQGEAQLAAYRLAVLRATED VENALTAVKREQQATVLAQVDALGKARGASAAAYEKGVVSLIEVLNADALLRTSD AQLQAQTEAARSAVAFAKALGGGWESTPATAVAVQEPAR"		
gene	5460..6083		
	/gene="amer"		
	/note="synonym: XCC0420"		
	5460..6083		
	/gene="amer"		
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"		
	/codon_start=1		
	/transl_table=11		
	/product="transcriptional regulator tetr/acrR family"		
	/protein_id="AAM39738.1"		
	/db_xref="GI:21111403"		
	/translation="MRNKAASSYPSVSGRSPADHDVDRDQIVNAATEHFRFRYGYEKTAVS DLAKISFYKAYIKFPEGAICEMICTNCLQIEDEVRAAVDETSPPEKFRMRMK VIVDASLFFEDRKLYEIAASAATERQSVLAYEERVLALLOEILQOGRQGGDFERK TELDEATRALVILIRPTNPNVLMQHSLDVIDEVLGSLGLVLSLSP"		
gene	6158..7273		
	/gene="mexE"		
	/note="synonym: XCC0421"		
	6158..7273		
	/gene="mexE"		
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"		
	/codon_start=1		
	/transl_table=11		
	/product="component of multidrug efflux system"		
	/protein_id="AAM39739.1"		
	/db_xref="GI:21111404"		
	/translation="MRRRLVVASLCLVPLALAGCGDKTPADPTATPLVRVATVGD ASSAARFSGTVAARVQSELGFRVSGSERLVDAQORVTRGQPLLRIQDVLKLAQ AQEAATVBARAQAGEDBARYDLRGTAISASAYDQIKAAADAKAQLSAQAAQ DVANANRYDLDLADQGVNEILVEFQVQVAVRVLAHAGREAILQLPETLOP VSGTAQATUFGNAGVTEPTELQLSDSADRLTRTFEARYVLGGALADAPLGTVSIR IPBAATQAGVQVPLGALFDAGKTGVWVTAGTFAKVAWVRPVRVLRDLDHNVAGTL ARGERIVAGLHLLRDGQHVRIAGVTAKPAPAAGQF"		
gene	7270..10368		
	/gene="XCC0422"		
	7270..10368		
	/gene="XCC0422"		
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"		
	/codon_start=1		
	/transl_table=11		



/protein id="AAM39740.1"  
/db\_xref="GI:21111405"  
/translation="MSEGRNLSALAVRERSITFLVLISLAGLVAFLKLRADPPA  
FTVKYNTIITAPGATPQEMQEVAEKLEKMQELRWDRITETIRPGGLATFTLLD  
STPGEVQGFQARKKVKYNIIGQPERIFVEFSDRLATLGVSPQDFVFAALNAQNA  
ADAETMRQLHVPQKVKYNIIGQPERIFVEFSDRLATLGVSPQDFVFAALNAQNA  
LTAAGSVTRGPDVFIIRDLGALNDLQKIRDTPLVQVGRSLKLSDIATVKRQYDPTF  
MIRSGEPALLGIIRMGWNGLDLGSDAEVAENAEPLGMQLSIVKVTQAVNIDA  
SVGEFTKFPFVALLVVLVCFVSMGRVGIIVAAAVPLTAAVFVVMILATCKNDRIT  
LQSLFALGLLVDDALIAEMVMKMEGGRVAAAVMSTAAAPMLKSGILVAVGF  
MNGFAASTAGEYTSNPKIIVGIALIVWAVVPTPLGVKMLPDMKIKEGGHAAY  
NTPYRNRQALGRVIARKLVAGSVGLFVLAVMGIVKQFFPIISDRPEVLVEVQ  
LPTGINSITAAAKSAMUSKQEKIIVTAVGQAPRFLFANGPELPDPSPFAKIV  
VTDQHOVMDKRLRAIAGLAPEARVATQLTFGYKSPFVAYRVSQDPTVIR  
GIAQVQVMDQSLRTNTDWTGVTPTLHFLSDQRLQAVGLTSAVAQLOFLIS  
GPIITLVREDIRSVQVARSAGDRLDPAIADFTLAGNGQVPLSOVGVKVDVME  
PVMRRDRVPTITCGVDVDTLQPDVSAAITROLOPIIDKLPSGVQIKKESIEEG  
KATAAMLPFPMLAATLIIILQVRSISAMVMFELTSLGLIGVPTLLIQQFPFGI  
NALVGLIALSGILMENTLIILOVRSISAMVMFELTSLGLIGVPTLLIQQFPFGI  
AFIPHTSVFNGTLATYLTGGLTGLVTLVFLPAMYSIWFKIRPDGKGNAPDIAR  
FA"

BASE COUNT 1668 a 3424 c 3575 g 1796 t

Alignment Scores:  
Pred. No.: 305 Length: 10463  
Score: 95.50 Matches: 38  
Percent Similarity: 38.48% Conservative: 17  
Best Local Similarity: 26.57% Mismatches: 46  
Query Match: 13.49% Indels: 42  
DB: 1 Gaps: 4

US-10-087-573-2 (1-141) x AE012139 (1-10463)

QY 26 ArgGluAlaLeuArgValLysSerGluArgLeuAlaMetLeuArgAlaLeuAla 45  
Db 2018 CGGCGCGGTTTGAGCGTGCACCCGCGCTACATCGGCTG-----GCCGTGGCA 2065  
QY 46 GlyMetCysGlyHisArgValLeuPro-----54  
Db 2066 GGGCGATCGTTGTGCGGTTGTGCCCCGACGATGACGTGTCGTGTCGTCGCGCT 2125  
QY 55 -----GlyThrGlyAlaSer-----59  
Db 2126 GACCATGTGAGCGTGGCGCTGAGCTTCCACGACGTGCGCCGCGCGCTGCTGGC 2185  
QY 60 -----AlaIleAlaAlaThrValThrPro-LysGlyAlaSerMetLy 73  
Db 2186 GATCTCGCGCGCGCGCTGGCAGCGCGCATCTACCGCAGCGCGCGCATTTGGA 2245  
QY 73 sLeuLysProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLy 93  
Db 2246 GCTGCGCGCGCATCTGCAACGAGCGCATCAAGGATGCGCGCGCTGTTGCGTGG 2305  
QY 93 sIleArgGluMetAenLysThrIleSerGlnGlnAlaArgValAenHisArgLeuPr 113  
Db 2306 CGTGGCCCACTGAGCGTGGTGGAGATGTCGGCGCGCTGTCGTCGTCGCGCC 2365  
QY 113 oGluGlyHisProLeu-----LeuGluLysArgAlaGluThrPheArgHi 128  
Db 2366 GCATCTGAGCGCGCTCTGGCAACGTTTGGAGCTAAGCAACGCGCGCGCTTTCACGCCA 2425  
QY 128 sLeuArg 130  
Db 2426 TTTGGCG 2432

RESULT 5  
HSM803142/c  
LOCUS  
DEFINITION Homo sapiens mRNA; cDNA DKFZp547P1613 (from clone DKFZp547P1613).  
ACCESSION AL031828  
VERSION AL031828.1 GI:21732302  
KEYWORDS  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 4656)  
AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Weil, B. and Wiemann, S.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUL-2002) 1. D-85764 Neuherberg, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by BMF (Biomedical Research Center at the Charite,  
Berlin/Germany) within the cDNA sequencing consortium of the German  
Genome Project.  
This clone (DKFZp547P1613) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: cloneszpd.de Further  
information about the clone and the sequencing project is available  
at http://mips.gsf.de/proj/cDNA/.  
FEATURES  
source  
1. 4656  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/clone="DKFZp547P1613"  
/tissue\_type="brain"  
/clone\_lib="547 (synonym: hfbr1). Vector pSport1; host  
DH10B; sites NotI + SalI"  
/dev\_stage="fetal"  
1. 4656  
/gene="DKFZp547P1613"  
3792. 4412  
/gene="DKFZp547P1613"  
/notes="similarity to cadherins"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD38541.1"  
/db\_xref="GI:21732303"  
/translation="MKGHVPPQLPVSTHPPSLPWTQBPSPCPSPQQTQLQPGPSH  
QLUTAPSAVKAPCPQPCPPASHPLCLLWLAHLPLHGLWPRVPPQSHSAQSLS  
HLMTRSCSTSGGPNATWDASHVFNAGPTLSLPWAASDYDLWLRGGHQAHLHP  
SLPLVLGPKQKCSLTCDQVQCKEYEGGSSPGFSSVRDPCPAPAP"  
4558. 4563  
/gene="DKFZp547P1613"  
4574  
/gene="DKFZp547P1613"  
BASE COUNT 953 a 1467 c 1307 g 929 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 215 Length: 4656  
Score: 92.50 Matches: 38  
Percent Similarity: 46.58% Conservative: 30  
Best Local Similarity: 26.03% Mismatches: 52  
Query Match: 13.06% Indels: 26  
DB: 9 Gaps: 6  
US-10-087-573-2 (1-141) x HSM803142 (1-4656)  
QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 3084 GCAAGCACCACCACTCCCAAGGAGGCTGCCACGCGCCCTACTAATAGC----- 3034  
QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 3033 -----AGTTGAGGTGAGGTGCCAGGCCAGTCCGCGGTGG 2998  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60  
Db 2997 GTGATATCACCGGTTCACAGGACCTGTGGCACTCCGGAACACGAGGAGGCCCCCGTGT 2938  
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGln 80  
Db 2937 ATCACCCTCCAGCTCAGCTCCGCTGCTTCCCGCGGTCCGCGCCCCCAGAG 2878

[illegible]

```

/codon_start=1
/transl_table=11
/protein_id="BAA75487.1"
/db_xref="GI:4514650"
/translation="MDAETTAEPVLPDTSAAAPGSGEAPTAPDAAEPGDDAAPSGAG
ATAAGPCGAAPVAPVAGVGRVEDFALYERAAALIHVYLLTGRGGDAPDSVEHAFOR
AWEHWPVEHDDPDVGVRSRAFEHALAPWERRPRPGPPADPVHRAALLRLPL
RTVLLGDGVCGLTVDEAAEAATVPATTSLEHARAALLERLPFVGPPEAREERLSA
LAGAVSTATPLAKSVTSGERRRLATGIVFGLTALLAGAVVFSAVTPRSPSPFAP
APYGSVDDGAPGGTSAAPGGGSPAPPSP"
4317.5237
/notes="see also D13898"
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="BAA75488.1"
/db_xref="GI:4514651"
/translation="MDRTTAPRNRRAVLATGVGAALATAAAGCPAHAPGRCARVEG
RLRALERTHAPALGAFAYDTGTGTAYRADERFPPIASMEFTIAVAVALRDLDQDE
LARRVHTADYKRSQSPVTGLPENVANGMTVAELCEATLTSNDNTAALLHDLGG
PVTATRCRSVGDHTRLDWEPELNSAEPGRVTDTSPTAIGRTYGRLLILGLLAAH
DRERLTRMLDNRKTSDBRFRKGLPADLLADKTTGGDYGTINNDAGVAMPGRPPVLA
VQTRFTPDAAEDNVLAEARLLAEMTD"

```

```

CDS
693 a 2192 c 2450 g 638 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 288 Length: 5973
Score: 92.50 Matches: 39
Percent Similarity: 39.44% Conservative: 17
Best Local Similarity: 27.46% Mismatches: 55
Query Match: 13.06% Indels: 31
DB: 1 Gaps: 4
US-10-087-573-2 (1-141) x AB024056 (1-5973)

```

```

QY 7 ThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu-ThrPheAspValMetAr 26
Db 1468 ACCGTGACGTGGTGGCTCTCGGAGCGGTACCGCAGGAGGACCGATCGGTACG 1409
QY 26 G-GluAlaLeuLeuArgValLys-----SerSerGluArgLeuA 39
Db 1408 ACGAGTTCCTCACGGGGTGGCTGAGCGGGCGAGTACTACGACGAGGAGCGCGC 1349
QY 39 laMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlas 59
Db 1348 ACATCACCGCGTGGGTGCTGGGGTGTCTGGCCACCGCGTCCGCGCGGAGCGCGC 1289
QY 59 erAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgP 79
Db 1288 AC-----CTGGCCCGCGTGTCTGGAGCGGCGCTCGCGAGCTGCTCGCCCGCGGTG 1235
QY 79 roGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle-----94
Db 1234 ACCGCGCGCGCGAGGCTTCGGTATCGAGGAGTTCACGCGCGCGCTCGCGGATGCGCACCG 1175
QY 95 -----ArgGluMetAsnLysThrIle-----101
Db 1174 GACCGGCTCTCGTACCGCGGAGTGGGACCGGACCGCGCTCTACACCGCTGGCGCGA 1115
QY 102 -----SerGluGluSerAlaArgValLeuHisArgLeuProGlyHisProLeuL 119
Db 1114 TGGTCCCGCGCGAGGAGTGTACCGGATCGTGGCGCGGCTCCCGCGCGTACCGCGCGC 1055
QY 119 eu 119
Db 1054 TC 1053

```

```

RESULT 8
AB027763/c
LOCUS
DEFINITION Streptomyces fradiae gene, partial cds.
ACCESSION AB027763

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS

```

```

AB027763.1 GI:4958945
Streptomyces fradiae
Streptomyces fradiae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 6573)
Ogawara,H.
Streptomyces
Published Only in DataBase (1999)
2 (bases 1 to 6573)
Ogawara,H.
Direct Submission
Submitted (26-MAY-1999) Hiroshi Ogawara, Meiji Pharmaceutical
University, Department of Biochemistry; Noshio,2-522-1, Kiyose,
Tokyo 204-8588, Japan (E-mail:hogawara@pharm.ac.jp,
Tel:81-424-95-8474, Fax:81-424-95-8474)
Location/Qualifiers
1.6573
/organism="Streptomyces fradiae"
/mol_type="genomic DNA"
/strain="Y53"
/db_xref="taxon:1906"
complement(<1..1310)
/notes="product is unknown
the 600th nucleotide of this sequence is corresponding to
the 1st nucleotide of the sequence of acc#AB024056"
/codon_start=1
/transl_table=11
/protein_id="BAA78104.1"
/db_xref="GI:4958946"
/translation="MARPADSAGVTGTPFAGTDAQVRGVFVDAHHNIMSNEAFGG
RLICGKPSERGVADALDCPEHYPDGSLALPDFITRGDGRHDPDGWTFADWPAHD
SLTHQONYVAVTBRALARGORVLNVLNTGVICSVYFFKDRGCDMESIRLQARTY
ENQEVYDMYMGGPGKWFRIUTDSQAAREVVRQGLAVLVVETSEPFQCKQYLDVAR
CEBIDRGLDLRLHRLGVRSMELCHKFNALCGVFDGALGTAVNVGGFLSTGTFWR
TPECTGPQNDPIGGASPAKLPVGVTPAYAGARCNTRGTLTDLGAYVRGNMR
ROMLEIDHMSVKAARALDILESEAYFGLSSHSMDMAEMTERVYRLGGFTIAQYMHG
SQAQTAERATDALARYGVGYGCTDMNMGVGPAPRGADAPDP"
BASE COUNT 770 a 2396 c 2685 g 722 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 323 Length: 6573
Score: 92.50 Matches: 39
Percent Similarity: 39.44% Conservative: 17
Best Local Similarity: 27.46% Mismatches: 55
Query Match: 13.06% Indels: 31
DB: 1 Gaps: 4
US-10-087-573-2 (1-141) x AB027763 (1-6573)
QY 7 ThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu-ThrPheAspValMetAr 26
Db 2068 ACCGTGACGTGGTGGCTCTCGGAGCGGTACCGCAGGAGGACCGATCGGTACG 2009
QY 26 G-GluAlaLeuLeuArgValLys-----SerSerGluArgLeuA 39
Db 2008 ACGAGTTCCTCACGGGGTGGCTGAGCGGGCGAGTACTACGACGAGGAGCGCGC 1949
QY 39 laMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlas 59
Db 1948 ACATCACCGCGTGGGTGCTGGGGTGTCTGGCCACCGCGTCCGCGCGGAGCGCGC 1889
QY 59 erAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgP 79
Db 1888 AC-----CTGGCCCGCGTGTCTGGAGCGGCTCGCGAGCTGCTCGCCCGCGGTG 1835
QY 79 roGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle-----94
Db 1834 ACCGCGCGCGCGAGGCTTCGGTATCGAGGAGTTCACGCGCGCGCTCGGGATGCGCACCG 1775
QY 95 -----ArgGluMetAsnLysThrIle-----101

```

```

Db      1774 GACCGGCTCTGTTACCGCGAGTGGACGACCGCCCTCTGACCAAGCTGGCGCGAGA 1715
Qy      102 -----SerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuL 119
Db      1714 TGGTGGCCGCGAGGAGTGGACGCGATCGTGGCGCGCTCCCGACGCGCTACGCGCGC 1655
Qy      119 eu 119
Db      1654 TC 1653

RESULT 9
AX367101/c
LOCUS      AX367101      10123 bp      DNA      linear      PAT 16-FEB-2002
DEFINITION Sequence 20 from Patent WO0198354.
ACCESSION AX367101
VERSION    AX367101.1 GI:18955303
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1
AUTHORS    Griffin,J.A., Kallick,D.A., Tribouley,C.M., Yue,H., Nguyen,D.B.,
            Tang,Y.T., Lal,P., Policky,J.L., Azimzai,Y., Lu,D.A., Graul,R.,
            Yao,M.G., Burford,N., Hatfield,A.J., Baughn,M.R., Bandman,O.,
            Patterson,C., Yang,J., Xu,Y., Warren,B.A., Ding,L. and
            Sanjanwala,M.S.
TITLE      Receptors
JOURNAL    Patent: WO 0198354-A 20 27-DEC-2001;
            Incyte Genomics, Inc. (US)
FEATURES   source
            1..10123
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /note="Incyte ID No: 1897612CB1"
BASE COUNT 1935 a 3207 c 2984 g 1997 t
ORIGIN
Alignment Scores:
Pred. No.: 537 Length: 10123
Score: 92.50 Matches: 38
Percent Similarity: 46.58% Conservative: 30
Best Local Similarity: 26.03% Mismatches: 52
Query Match: 13.06% Indels: 26
DB: 6 Gaps: 6

US-10-087-573-2 (1-141) x AX367101 (1-10123)
Qy      3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
Db      8612 GCAAGCACCACTCCCAAGGAGGTGCCAGGCGCCCTACTAATAGC----- 8562
Qy      23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
Db      8561 -----AGGTTGAGGTGAGGTGCCAGGCGCCAGTCCGCGGTGG 8526
Qy      43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60
Db      8525 GTGATATCCAGGTACAGGACTGTGGCACTCCGGACACGAGGAGGCGCCCGTGTCT 8466
Qy      61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80
Db      8465 ATCACTCCAGGCTCAGCTCCCGTGTGTCTCCGCGCGGTCCGCGCGCCCGCCAGAG 8406
Qy      81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95
Db      8405 GTGGCTGTTCCGCTGCTGTCGCGCACTGTCCACCGCAGATACAGGCTCTCTAGTC 8346
Qy      96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisAr 111
Db      8345 TGGTTAATACCAATAATGGGGAAGAGGTGGCAAGGAA---TACAGAACCAGGCGCATCG 8289

```

```

Qy      111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArgHisLe 129
Db      8288 GCACCCCATCTCTCATCTGTGGCTGCACGTGACCAAGCTGGCCACCGCGGCACCT 8229
Qy      129 uArgSerLeuLysSer 134
Db      8228 TCGGGCACCTTGGAGT 8213

RESULT 10
BD084093/c
LOCUS      BD084093      10759 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION Novel human datus gene and protein encoded thereby.
ACCESSION BD084093
VERSION    BD084093.1 GI:22629703
KEYWORDS   JP 2001327295-A/1.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1
AUTHORS    Nakayama,M., Nagase,T., Ohara,O. and Nakajima,D.
TITLE      Novel human datus gene and protein encoded thereby
JOURNAL    Patent: JP 2001327295-A 1 27-NOV-2001;
            KAZUSA DNA RESEARCH INSTITUTE
COMMENT     OS Homo sapiens (human)
            PN JP 2001327295-A/1
            PD 27-NOV-2001
            PF 27-JUL-2000 JP 2000226291
            PI MANABU NAKAYAMA, TAKAHIRO NAGASE, OSAMU OHARA, DAISUKE NAKAJIMA
            PC C12N15/09,C07K14/47//A61K31/711.A61K38/00,A61K39/395,A61K48/00, PC
            A61P43/00,
            PC C12N15/00,A61K37/02
            CC Novel human datus gene and protein encoded thereby FH Key
            Location/Qualifiers
            FT CDS
            source
            1..10759
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
BASE COUNT 2008 a 3412 c 3256 g 2083 t
ORIGIN
Alignment Scores:
Pred. No.: 577 Length: 10759
Score: 92.50 Matches: 38
Percent Similarity: 46.58% Conservative: 30
Best Local Similarity: 26.03% Mismatches: 52
Query Match: 13.06% Indels: 26
DB: 6 Gaps: 6

US-10-087-573-2 (1-141) x BD084093 (1-10759)
Qy      3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
Db      9265 GCAAGCACCACTCCCAAGGAGGTGCCAGGCGCCCTACTAATAGC----- 9215
Qy      23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
Db      9214 -----AGGTTGAGGTGAGGTGCCAGGCGCCAGTCCGCGGTGG 9179
Qy      43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60
Db      9178 GTGATATCCAGGTACAGGACTGTGGCACTCCGGAAACGAGGAGGCGCCCGTGTCT 9119
Qy      61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80
Db      9118 ATCACTCCAGGCTCAGCTCCCGTGTGTCTCCGCGCGGTCCGCGCGCCCGCCAGAG 9059
Qy      81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95

```



**TITLE** Direct Submission  
**JOURNAL** Submitted (27-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Mar 14, 2001 this sequence version replaced gi:12958053.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997).  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: rz162  
 Center clone name: 211 E 17

```

FEATURES             Location/Qualifiers
     1..192239
     organism=Homo sapiens"
     molecule_type="genomic DNA"
     db_xref="taxon:9606"
     chromosome="11"
     map="11"
     clone="RP11-211E17"

```

```

repeat_region 2608. .2856  

/rpt family="MIR"  

complement(2943. .3155)  

/rpt family="L2"  

repeat_region complement(3383. .3472)  

/rpt family="MIR"  

repeat_region complement(3457. .4176)  

/rpt family="L2"  

repeat_region 4598. .4665  

/rpt family="L2"  

repeat_region complement(4670. .4721)  

/rpt family="MIR"  

repeat_region complement(4741. .5054)  

/rpt family="AluSx"  

repeat_region 5081. .5535  

/rpt family="L2"  

repeat_region 5596. .6147  

/rpt family="L2"  

repeat_region 6166. .6246  

/rpt family="MIR"  

repeat_region complement(6247. .6562)  

/rpt family="AluJb"  

repeat_region 6563. .6626  

/rpt family="MIR"  

repeat_region 6628. .6758  

/rpt family="FLAM_A"  

repeat_region 6776. .7085  

/rpt family="AluSg"  

repeat_region 7088. .7135  

/rpt family="AT_rich"  

repeat_region 7136. .7247  

/rpt family="MIR"  

repeat_region complement(7544. .7681)  

/rpt family="L2"  

repeat_region complement(7794. .7875)  

/rpt family="L2"  

repeat_region 7977. .8441  

/rpt family="L2"  

repeat_region 8442. .8546  

/rpt family="L1MA4A"  

repeat_region 8547. .8561  

/rpt family="L2"  

repeat_region 8562. .8862  

/rpt family="AluSx"  

repeat_region 8863. .8970  

/rpt family="L2"  

repeat_region 9753. .9887  

/rpt family="MIR"

```

repeat_region	10142..10678
repeat_region	/rpt_family="L2"
repeat_region	10772..10802
repeat_region	/rpt_family="AT rich"
repeat_region	complement(10875..11336)
repeat_region	/rpt_family="L2"
repeat_region	11337..11632
repeat_region	/rpt_family="AluJo"
repeat_region	complement(11633..11854)
repeat_region	/rpt_family="L2"
repeat_region	11855..11987
repeat_region	/rpt_family="FLAM_C"
repeat_region	complement(11988..12619)
repeat_region	/rpt_family="L2"
repeat_region	12708..12749
repeat_region	/rpt_family="(TGAA)n"
repeat_region	complement(13509..13581)
repeat_region	/rpt_family="L2"
repeat_region	complement(14255..14430)
repeat_region	/rpt_family="FRAM"
repeat_region	complement(14431..14945)
repeat_region	/rpt_family="L2"
repeat_region	complement(15389..15675)
repeat_region	/rpt_family="L2"
repeat_region	15930..15862
repeat_region	/rpt_family="(TAAA)n"
repeat_region	15964..15988
repeat_region	/rpt_family="L2"
repeat_region	15997..16125
repeat_region	/rpt_family="MIR3"
repeat_region	16453..16522
repeat_region	/rpt_family="L2"
repeat_region	16579..16716
repeat_region	/rpt_family="L3"
repeat_region	17220..17269
repeat_region	/rpt_family="G-rich"
repeat_region	17729..17789
repeat_region	/rpt_family="GC-rich"
repeat_region	17870..17966
repeat_region	/rpt_family="C-rich"
repeat_region	complement(18511..18687)
repeat_region	/rpt_family="MIR"
repeat_region	19387..19540
repeat_region	/rpt_family="L2"
repeat_region	complement(19541..19849)
repeat_region	/rpt_family="AluY"
repeat_region	19850..20110
repeat_region	/rpt_family="L2"
repeat_region	complement(24563..24667)
repeat_region	/rpt_family="L2"
repeat_region	complement(26418..26529)
repeat_region	/rpt_family="MIR"
repeat_region	complement(27770..27888)
repeat_region	/rpt_family="MIR"
repeat_region	complement(27951..28252)
repeat_region	/rpt_family="AluSg"
repeat_region	complement(28565..28634)
repeat_region	/rpt_family="MIR"
repeat_region	29114..29357
repeat_region	/rpt_family="L2"
repeat_region	29497..30535
repeat_region	/rpt_family="L2"
repeat_region	30607..31076
repeat_region	/rpt_family="L2"
repeat_region	complement(32339..32599)
repeat_region	/rpt_family="THE1B"
repeat_region	33353..33456
repeat_region	/rpt_family="L2"
repeat_region	complement(33788..34079)
repeat_region	/rpt_family="AluSg"
repeat_region	34150..34182
repeat_region	/rpt_family="AT rich"
repeat_region	complement(34350..34646)

Alignment Scores:  
Pred. No.: 1.72e+04 Length: 192239  
Score: 92.50 Matches: 38  
Percent Similarity: 46.58% Conservative: 30  
Best Local Similarity: 26.03% Mismatches: 52  
Query Watch: 13.06% Indels: 26  
DB: 9 Gaps: 6

US-10-087-573-2 (1-141) x AC009796 (1-192239)

Qy	3	SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyIuThrPhe	22
Db	16656	GCAAGCACCACCACTCCCAAGGAGGCTGCCAGCGCCCTACTAATAGC-----	166706
Qy	23	AspValMetargGluAlaLeuLeuargVallysSerSerGluArqLeuAlametLeuarg	42
Db	166707	-----AGGTGGAGTCAGTGCAGCGCCAGCGCGCTGTGGTGGG	166742
Qy	43	AlaLeuAlaGlyMetCysGlyHisArgValleu-----ProGlyThrGlyAlaSerAla	60
Db	166743	GTGATATCCACGGTCACAGGCACTGTGGCACTCCCGGAACAGGCAGAGCCCCCGTGCT	166802
Qy	61	IleAlaIalatrValThrProLysGlyAlaSerMetLysLeuLysProIarGln	80
Db	166803	ATCACCTCCAGCCTCAGTCCCGTGGTCTTCCCGCGGGTCCCGGCCCCACCCACAG	166862
Qy	81	--SerThrLysSerProGluLeuArggluLeuSerArgLysIleArg-----	95
Db	166863	GTGGCTGTTCGGCTGCCTGGTGGCCGACTGTCCACCCGAGGTACAGGCTCCTGTAGTC	166922
Qy	96	-----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal--AsnHisAr	111
Db	16923	TGGTTAATACAAATAGGGGAGAGAGGTGGCAAGGAA---TACAGAACCAGGCCATCG	169799
Qy	111	gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrrPheArgHisLe	129
Db	16980	GCACCCCATTCTCTGTGGCTTGACGTGACCCAGTGTGGCCAGCGCGGCACCT	167039
Qy	129	uArgSerLeuLysSer 134	
LJ	167040	TCGGCAGCTTGGAGT 167055	
RESULT 13			
AC027641/c			
LOCUS	AC027641	208430 bp DNA linear HTG 07-JUL-2000	
DEFINITION	Homo sapiens chromosome 11 clone RP11-732A19, WORKING DRAFT		
SEQUENCE	SEQUENCE, 18 unordered pieces.		
AC027641			
ACCESSION	AC027641.2 GI:8570385		
VERSION	HTG; HTGS PHASE1; HTGS_DRAFT.		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	Waterston,R.H. The sequence of Homo sapiens clone		
AUTHORS	Unpublished		
JOURNAL	Waterston,R.H. 2 (bases 1 to 208430)		
REFERENCE	Direct Submission		
AUTHORS	Submitted (30-MAR-2000) Genome Sequencing Center, Washington		
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
COMMENT	On Jun 17, 2000 this sequence version replaced gi.7344780.		
	----- Genome Center -----		
	Center: Washington University Genome Sequencing Center		
	Center code: WUGSC		
	Web site:http://genome.wustl.edu/gsc/index.shtml		
	----- Project Information -----		
	Center project name: H NH0732A19		



```
----- Summary Statistics -----
Sequencing vector: M13; 98%
Chemistry: Dye-terminator ET; 98% of reads
Chemistry: Dye-terminator Big Dye; 2% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 198825 bases at least Q40
Consensus quality: 202103 bases at least Q30
Consensus quality: 203934 bases at least Q20
Insert size: 219000; agarose-fp
Quality coverage: 4.27 in Q20 bases; agarose-fp
Quality coverage: 4.55 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1760: contig of 1760 bp in length
* 1761 1860: gap of unknown length
* 1861 3171: contig of 1311 bp in length
* 3171 3271: gap of unknown length
* 3271 5628: contig of 2357 bp in length
* 5628 5729: gap of unknown length
* 5729 7760: contig of 2032 bp in length
* 7760 7861: gap of unknown length
* 7861 15503: contig of 7643 bp in length
* 15503 15604: gap of unknown length
* 15604 18871: contig of 3268 bp in length
* 18871 22431: gap of unknown length
* 22431 22530: contig of 3459 bp in length
* 22530 22536: gap of unknown length
* 22536 28557: contig of 6026 bp in length
* 28557 28557: gap of unknown length
* 28557 35414: contig of 6758 bp in length
* 35414 35515: gap of unknown length
* 35515 42738: contig of 7224 bp in length
* 42738 42739: gap of unknown length
* 42739 50307: contig of 7469 bp in length
* 50307 50407: gap of unknown length
* 50407 62974: contig of 12567 bp in length
* 62974 63074: gap of unknown length
* 63074 73754: contig of 10679 bp in length
* 73754 73853: gap of unknown length
* 73853 7730: contig of 13777 bp in length
* 7730 87631: gap of unknown length
* 87631 106802: contig of 19072 bp in length
* 106802 106902: gap of unknown length
* 106902 120751: contig of 13748 bp in length
* 120751 120750: gap of unknown length
* 120750 156637: contig of 35887 bp in length
* 156637 156738: gap of unknown length
* 156738 208430: contig of 51693 bp in length.
*
* Location/Qualifiers
* 1..208430
* /organism="Homo sapiens"
* /mol_type="genomic DNA"
* /db_xref="taxon:9606"
* /chromosome="11"
* /clone="RP11-732A19"
*
* 1..1760
* /note="assembly_name:Contig8"
* 1861..3171
* /note="assembly_name:Contig9"
* 3271..5628
* /note="assembly_name:Contig10"
* 5729..7760
* /note="assembly_name:Contig11"
* 7861..15503
```

```
misc_feature /note="assembly_name:Contig12"
15604..18871
/note="assembly_name:Contig13
clone_end:T7
vector_side:right"
18972..22430
/note="assembly_name:Contig14"
22531..28556
/note="assembly_name:Contig15"
28657..35414
/note="assembly_name:Contig16"
35515..42738
/note="assembly_name:Contig17"
42839..50307
/note="assembly_name:Contig18"
50408..62974
/note="assembly_name:Contig19"
63075..73753
/note="assembly_name:Contig20"
73854..87630
/note="assembly_name:Contig21"
87731..106802
/note="assembly_name:Contig22"
106903..120650
/note="assembly_name:Contig23
clone_end:SP6
vector_side:left"
120751..156637
/note="assembly_name:Contig24"
156738..208430
/note="assembly_name:Contig25"
BASE COUNT 59749 a 44894 c 43435 g 58647 t 1705 others
ORIGIN
Alignment Scores:
Pred. No.: 1.89e+04 Length: 208430
Score: 92.50 Matches: 38
Percent Similarity: 46.58% Conservative: 30
Best Local Similarity: 26.03% Mismatches: 52
Query Match: 13.06% Indels: 26
DB: 2 Gaps: 6
US-10-087-573-2 (1-141) x AC027641 (1-208430)
QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
Db 153648 GCAAGCACCAACCAACTCCCAAGGAGGCTGCCACGCGCCCTACTAATAGC----- 153598
QY 23 AspValMetArgGluAlaLeuLeuArgVallyssSerGluArgLeuAlaMetLeuArg 42
Db 153597 -----AGGTTGAGGTCAGGTGCCAGGCCAGTGGCGGTGGT 153562
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60
Db 153561 GTGATATCCACGGTCACAGGCACCTGTGGCACTCCGGGAACCAAGCAGAGGCCCGGTGCT 153502
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80
Db 153501 ATCACCCTCCAGCTCAGCTCCCGGTGCTTCCCGCGCGGTGCCGCGCCACCCAGAG 153442
QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95
Db 153441 GTGGCTGTTCGGCTGCTGTGGTCCCGACTGTCCACCCGCGAGTACAGGCTCCTGTAGTC 153382
QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal--AsnHisAr 111
Db 153381 TGGTTAATACCAAAATAGGGGAGAGAGGTGGCAAGGAA---TACAGAACCCAGCCATCG 153325
QY 111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluThrPheArgHisLe 129
Db 153324 GCACCCCATCTCTATCTGTGGCTGTGCACGTGACCCCAAGCTGTGGCCACGCGGGCACCT 153265
QY 129 uArgSerLeuLysSer 134
```





gene	CDS
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

```

GHLSPVAFERHLLFTLGGWRRKEELDRSVNTAGDSARHKAFLQDEIAFGDWSLTL
GRFDKHEAFGRSSFRLLYLLHLLSDVLTLAGRGVGRYKAPSKQLSPSYAAGGGGR
FTIYNPDLPKPTNTSYELGADYQDGMKSLGTFTFENDVRLDIQTVVARCGVGGEI
RVNENVDRIIRGLELGGGVDPADLRWELNVTYDARNRTAGORLGRSRHLANSOL
RASPVRFSQAQLRTEVVGSOVAYSSNVGVALPAYSLWHLSELSKLSNLTLRGGIENL
GGRLADDTHETIYAEPRGTFLHLGLVASF"
6136..9981
/locus_tag="PA1923"
6136..9981
/locus_tag="PA1923"
/notes="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG05311.1"
/db_xref="GI:9947917"
/transl_table="MRSLLFCGALIASCAWPAQVRLSSDFVLPGKHORLAGWA
RAGVRLRLGRLIGIGAPPEWLDGNNLLILOTPTDRAQVEEALGERLGGTGPWI
RVGGPGPGNIPALGRLGVYANGSEANRLRLEFVVRWAGLTVDALPAPOLA
QAGFYHDPAPAVLADYLANGASWASDAPRIAFILIPRGALADAQTADELERS
ERHQAPLAWFDDSDPEALRKSFGADQVALVNIQHQNGPARRAEALDVPVLOT
LGRDNEADLWLAASGAPRTAAAFGLMPETWGMSSDPLVISALENGEFLKAWGAE
LJDKLRLRLRLPAADKHALMFNWHPEGEKNVAASHLNPASLRLGEALRAAG
RVATSDSALIDTAQLCGYRPTDLDALYRDGLAASLPDYLHFWFEALPADREE
MRWMDPDRHWRALRIDGRRFVFPAAALGNLLLPQPRACRGEAVHDSAPVPH
LVLAVYQVREGGADVLHFGHTQEWLPGKDRGLAVGYELRALGDLVPFYVIO
DNVGEAIVARRGGVATVSHQTPSPAPAGLYDELRLHQLIHEYQQLDEGAVRSZ
QIRAVRAHNNDDLGNSAMHEDPPAFGLVHLHLRKLASMPGLHTIPGVAASP
EURLGTVMQOGEFYRAJGLDPELFAADFRALREGRAYRTLQRYLRDGGBIKAD
PRLQELRLARLDRLGADTGLLEALLAGLRAFGVAPGGDPINRQVPSGRNLFAP
EADKVTRAAVYAGAFQQLLESYAEHQRAPEKLAFLWSSETMRHLGVESQAL
HALGLPRWDAGRLIALDIVPAALRGPRVDVLTQVTSVYRDQDFGFMRLIAEIER
LAALDPGNLARNQALERRRLEVEAPLQRLSLRFLGNAPGDYGTGTQTLTLD
STWDDSDSALQFLQYAVSGSDMGVKLQDGNLFAQLKGVQAAILLRSSTLNG
LSTDHPEVYGLSLAVRHLHGASPALYIADLRQRPRTAAQFLASELRGRYLSPO
WIAHQRETAGSLEMLDANNLWQADRLMRADQWQALHDTFVMDRRLGUAEF
FETHNTAQAIIRAEAIRKYWDASEQTRRELAERWRQLAAGAEVGAATREF
IERMAGSGFEAAAGAAAGAGETVSGRVLEEVAPPSPGPPQPLRLAARALLFLF
AAGAVRQSVNRRPMENRS"
9978..10454
/locus_tag="PA1924"
9978..10454
/locus_tag="PA1924"
/notes="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11

Alignment Scores:
Pred. No.: 659 Length: 11060
Score: 92.00 Matches: 44
Percent Similarity: 46.03% Conservative: 14
Best Local Similarity: 34.92% Mismatches: 51
Query Match: 12.99% Indels: 17
DB: 1 Gaps: 5

US-10-087-573-2 (1-141) x AE004618 (1-11060)

Qy 25 MetArgGluAlaLeuArgValLysSerSerGluArg---LeuAla----- 39
Db 8362 CTGCGGAGCAGTGTGTCGCGCGCGCACTCGACCGCAGCTCGCGCACCGCGAA 8421
Qy 40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59
Db 8422 CTGAGGCGTGTCTCGCGCGCTTCGCGCGCGCTTCGTCGCGCGCGCGCGCGAT 8481
Qy 60 Alaile-AlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgPr 79
Db 8482 CCGATCCGCAATCCGCGAGGTACCCAGCGCGCGCAACC-----TGTTCGCTTCGAGGCC 8535

```

```

Qy 79 o-GlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg-GluMetAsn 98
Db 8536 GACAAGGTACCGACCGCGCGCTACGAAGCCGCGCGCTTCGCGCAACTG--- 8592
Qy 99 LysThrIleSerGlnGluSerAlaArgValAsnHis-----ArgLeuProGluGlyHis 116
Db 8593 -----CTGGAAGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8640
Qy 117 ProLeuLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGly 136
Db 8641 GCCTTCAGCCTCTCGTCTCGGAACCATGCCACCTCGGCGCATCGTCGAAGCCAGCGC 8700
Qy 137 ValAsnArgLeu 140
Db 8701 CTGACGCGCCCTC 8712

RESULT 17
AF416730/c
LOCUS AF416730 1386 bp mRNA linear ROD 25-OCT-2001
DEFINITION Rattus norvegicus CORO1A protein mRNA, complete cds.
ACCESSION AF416730
VERSION AF416730.1 GI:16417359
KEYWORDS SOURCE
ORGANISM Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1386)
AUTHORS Suzuki,K., Takeshita,F., Nakata,N. and Makino,M.
TITLE Molecular cloning of rat Corola
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1386)
AUTHORS Suzuki,K., Takeshita,F., Nakata,N. and Makino,M.
TITLE Direct Submission
JOURNAL Submitted (05-SEP-2001) Department of Microbiology, Leprosy
Research Center, National Institute of Infectious Diseases, 4-2-1
Aoba-cho, Higashimurayama, Tokyo 189-0002, Japan
FEATURES
source
1..1386
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="spleen"
1..1386
/codon_start=1
/product="CORO1A protein"
/protein_id="AA118695.1"
/db_xref="GI:16417360"
/translation="MSRQVRSKFRHVFQPAKADOCYEDVRVSVQTTWDSGFCVAVNP
KFVALICEASGGGAFVLPGLKTRGVKNVPLVCGHTAPVLDIAWCPHNDNVIASGSE
DCTVMWEIPDGLVPLRFPVTVLEGTTRGVIAWHPTAQNVLLSAGCDNVILVWD
EGTRPVHAFVSEGLITTTGFSRMSRQVALWTKHLEELSLDELDTSSGVLLPFFD
PDTNIVYLCGKGSSIRYFITSAPFLHYLSMESSESORSGMGYMPKRLGVNKKCI
ARFKYLERKCEPIATVPKSDLFQEDLFPPTAGPPDPAITAEWLSDRAGPLLLSL
KGVFPVPSKRELVRNGLDSARRKATPEPSTLSSDTSVRLEEDVRNLNAILVQLQR
LDRLEETVQAK"
BASE COUNT 291 a 373 c 429 g 293 t
ORIGIN
Alignment Scores:
Pred. No.: 63.2 Length: 1386
Score: 91.50 Matches: 35
Percent Similarity: 42.61% Conservative: 14
Best Local Similarity: 30.43% Mismatches: 31
Query Match: 12.92% Indels: 35
DB: 10 Gaps: 6
US-10-087-573-2 (1-141) x AF416730 (1-1386)

```



Db 264 CCCCCCA-----CTGGCCTCACAGATC 241  
 Qy 95 ArgGluMetAsn-----LysThrIleSerGln 103  
 Db 240 AGACCCATGAACCTGGGTTGACAGCGCAGAGCCACTGTCCCAA 196  
 RESULT 19  
 AF495469/c 1618 bp mRNA linear ROD 23-APR-2002  
 LOCUS Rattus norvegicus tryptophane aspartate-containing coat protein  
 DEFINITION (Taco) mRNA, complete cds.  
 ACCESSION AF495469  
 VERSION AF495469.1 GI:20271116  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 1618)  
 Kohchi,C., Inagawa,H., Makino,K., Terada,H. and Soma,G.-I.  
 A new therapeutic strategy of mycobacterium infection by use of  
 anti-TACO sequence  
 Unpublished  
 2 (bases 1 to 1618)  
 Kohchi,C., Inagawa,H., Makino,K., Terada,H. and Soma,G.-I.  
 Direct Submission  
 TITLE Submitted (26-MAR-2002) Radioisotope Center, Hiroshima University,  
 1-4-2 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan  
 JOURNAL Location/Qualifiers  
 FEATURES  
 source  
 1. .1618  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="F34"  
 /db\_xref="taxon:10116"  
 /sex="female"  
 /tissue\_type="brain"  
 1. .1618  
 /gene="Taco"  
 105. .1490  
 /gene="Taco"  
 /codon\_start=1  
 /product="tryptophane aspartate-containing coat protein"  
 /protein\_id="AA18515.1"  
 /db\_xref="GI:20271117"  
 /translation="MSRQVRSKFRHVFQPAKQCYEDVRVSQTTWDSGFCVNP  
 KFMALICEAGSGGAFVLPLGKTRVDKNVPLVCGHTAPVLDIAWCPHNDNVIASGE  
 DCTVMWEIPDGLVPLPRPVVTLGHTKRVGIWAHPPTAQNVLISAGCDNVILVMD  
 VGTGAALTLGPDVHPDTIYSVDSMRDGLICTSCRDKRVVLEPKGTVAEKERPH  
 EGPVHAVFVSGKILTTGFSMSRQVALWDTKHEPLSLQELDTSSGVLLPFFD  
 PDNIYVLCGSDSIIRYFITSAPFLHYLSMFSKESQRMGMKPRGLEVNKCEI  
 ARFYKHERKCEPIAMTVPRKSLDFQEDLYPTAGPDPAALAEWLSGRDAGPLLI  
 LDRLLEETVQAK"  
 BASE COUNT 353 a 444 c 476 g 345 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 75.8 Length: 1618  
 Score: 91.50 Matches: 35  
 Percent Similarity: 42.61% Conservative: 14  
 Best Local Similarity: 30.43% Mismatches: 31  
 Query Match: 12.92% Indels: 35  
 DB: 10 Gaps: 6  
 US-10-087-573-2 (1-141) x AF495469 (1-1618)  
 Qy 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 487 TCCAAAGGTGACACAGGCTCCCGAGCGCAGCAGCCACCCCTCCGAGATCTCCAC 428  
 Qy 23 AspValMetArgGluAlaLeuLeuArgValIleSerSerGluArgGluAlaMetLeuArg 42

Db 427 ACCATA-----ACTGTGAGTCTCTCAGAGCCACTGGCAATGACGTTG 386  
 Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyVala----- 58  
 Db 385 TCATTG-----TGGGACACAGCGGATGCTTACACAGGGGCGAGTGTGGCCGAC 335  
 Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 Db 334 ACCAGGGGACCATCTTCTACTCTGCTCAGTCTTGCCTAGGGGTAGCACCAGGAAGGCC 275  
 Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 Db 274 CCTCCCCCA-----CTGGCCTCACAGATC 251  
 Qy 95 ArgGluMetAsn-----LysThrIleSerGln 103  
 Db 250 AGAGCCATGAACCTGGGTTGACAGCGCAGAGCCACTGTCCCAA 206  
 RESULT 20  
 AF143955/c 1633 bp mRNA linear ROD 27-MAY-1999  
 LOCUS Mus musculus coronin-1 mRNA, complete cds.  
 DEFINITION AF143955  
 ACCESSION AF143955.1 GI:4895036  
 VERSION  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Okumura,M., Kung,C., Wong,S., Rodgers,M. and Thomas,M.L.  
 Definition of family of coronin-related proteins conserved between  
 humans and mice; close genetic linkage between coronin-2 and  
 CD45-associated protein  
 JOURNAL DNA Cell Biol. 17 (9), 779-787 (1998)  
 MEDLINE 98449467  
 PUBMED 9778037  
 REFERENCE 2 (bases 1 to 1633)  
 Okumura,M., Kung,C.E., Wong,S., Rodgers,M. and Thomas,M.L.  
 Direct Submission  
 TITLE Submitted (19-APR-1999) Pathology, Washington University School of  
 Medicine, 660 South Euclid Avenue, St. Louis, MO 63110, USA  
 JOURNAL Location/Qualifiers  
 FEATURES  
 source  
 1. .1633  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 138. .1523  
 /codon\_start=1  
 /product="coronin-1"  
 /protein\_id="AAD32703.1"  
 /db\_xref="GI:4895037"  
 /translation="MSRQVRSKFRHVFQPAKQCYEDVRVSQTTWDSGFCVNP  
 KFMALICEAGSGGAFVLPLGKTRVDKNVPLVCGHTAPVLDIAWCPHNDNVIASGE  
 DCTVMWEIPDGLVPLPRPVVTLGHTKRVGIWAHPPTAQNVLISAGCDNVILVMD  
 VGTGAALTLGPDVHPDTIYSVDSMRDGLICTSCRDKRVVLEPKGTVAEKERPH  
 EGPVHAVFVSGKILTTGFSMSRQVALWDTKHEPLSLQELDTSSGVLLPFFD  
 PDNIYVLCGSDSIIRYFITSAPFLHYLSMFSKESQRMGMKPRGLEVNKCEI  
 ARFYKHERKCEPIAMTVPRKSLDFQEDLYPTAGPDPAALAEWLSGRDAGPLLI  
 LDRLLEETVQAK"  
 BASE COUNT 352 a 447 c 496 g 338 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 76.6 Length: 1633  
 Score: 91.50 Matches: 35  
 Percent Similarity: 42.61% Conservative: 14  
 Best Local Similarity: 30.43% Mismatches: 31  
 Query Match: 12.92% Indels: 35  
 DB: 10 Gaps: 6  
 US-10-087-573-2 (1-141) x AF143955 (1-1633)



```

Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 520 TCAAGGTGACGACAGGCTCCCGAGCGGACGACCGCCCATCGGATCTCCAC 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 460 ACCATA-----ACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTG 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 TCATTG-----TGTGGACACCGCGGATGTACACAGGGCGAGTGTGGCGGAG 368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 367 ACCAGGGGCGACGCTTCTGCTACTCGTCCAGTCTTGCCTAGGGTAGCACAGGAAGGCC 308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 307 CTTCCCCCA-----CTGGCTTCACAGATC 284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 95 ArgGluMetAsn-----LysThrIleSerGln 103
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 283 AGAGCCATGAACCTTGGGTGTGACAGCGAGGACCACTGTCCAA 239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 21
HSCORONIN/c HSCORONIN 1563 bp mRNA linear PRI 23-DEC-1995
LOCUS
DEFINITION H.sapiens mRNA for coronin.
ACCESSION X89109
VERSION X89109.1 GI:1136139
KEYWORDS coronin homologue.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Grogan,A., Keep,N.H., Reeves,E. and Segal,A.W.
JOURNAL Unpublished
PUBLISHED
2 (bases 1 to 1563)
REFERENCE
AUTHORS Keep,N.H.
JOURNAL Direct Submission
TITLE Submitted (23-JUN-1995) N.H. Keep, University College London,
Medicine, Rayne Institute, 5 University Street, London, WC1E 6JF,
UK
FEATURES
Location/Qualifiers
1..1563
/organism="Homo sapiens"
/mol_type="mRNA"
/strain="ATCC 8531"
/db_xref="taxon:9606"
/tissue_type="brain"
/dev_stage="infant"
/lab_host="E.coli"
94..1479
/codon_start=1
/product="coronin homologue"
/protein_id="CAA61482.1"
/db_xref="GI:1136140"
/db_xref="SWISS-PROT:P31146"
/translation="MSRQVRSKPRHVFQGPAPAKADQCYEDVRVSQTTWDSGFCVAVP
KEVALICASGGAGFLVPLGKGTVDKNAPTVCGHTAPVLDIAWCHNDNVIASGSE
DCTVMWEIPDGLMLPREPVVTEGHTKRGIVAMHTTAQNVLSSAGCDNVLMVWD
VTGAAMTLGPEVHPDTIYSVDWSRDGLICTSCRDKRVRIIEPRKGTVAEKDRPH
EGTRPVRAVFSVEGKLTITGFSRMSERQVALWDTKHEEPLSLQELDTSSGVLLPFD
PDNTNYLCGGKDSIRYFEITSEAPFLHYSMFSSKESQRMGMXPXRGVLEVNKEI
ARFYKHERCEPIAMTVPRKSDLFQEDLYPTTAPGDPALTAEEWLGDRDAGPLLISL
KDGVPKPSRELKRVNRGLDTGRRRAAPEASGTPSSDAVSRLLEEEMRKQLQATVQELQK
LDRLKETVQAK"
300 a 492 c 494 g 277 t
BASE COUNT
ORIGIN
Alignment Scores:

```

```

Pred. No.: 98.4 Length: 1563
Score: 90.00 Matches: 40
Percent Similarity: 41.54% Conservative: 14
Best Local Similarity: 30.77% Mismatches: 41
Query Match: 12.71% Indels: 35
DB: 9 Gaps: 6

US-10-087-573-2 (1-141) x HSCORONIN (1-1563)

Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 494 ACAGCTTGTGTGGCCCTCCAGGTTGACAGCGGCTCCCGAGGGGACGATCAGGCC 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 434 CCATCGGGATCTCCACACCATGACTGTGTCAGTCTCGGAGCCACTGGCAATGACGTTG 375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 374 TCATTG-----TGTGGGACACCGCGGATGTCTAGCACAGGGGCTGTGTGGCCACAG 324
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 323 ACCGTGGGCGCATCTTGTTCACACAGCTCCAGTCTTGCACAGGGGACGACAGGAAGGCC 264
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 263 CTTCCCCCG-----CTGGCTTCACAGATC 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 AGGCCCAAACTTAGGGTTGACAGCACAGACCACTGTCCAGGTTGGTCTGTGAGACG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 108 LAsnHisArgLeuProGluGlyHisPro 117
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 CGCACATCTTCATAGCACTGTGTGGGCT 152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 22
HSU34690/c HSU34690 1589 bp mRNA linear PRI 02-OCT-1995
LOCUS
DEFINITION Human coronin-like protein (HCO001) mRNA, complete cds.
ACCESSION U34690
VERSION U34690.1 GI:1002922
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1589)
REFERENCE
AUTHORS Liau,G., Popa,I., Argraves,K. and Argraves,W.S.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-1995) Gene Liau, Dept. of Molecular Biology,
Holland Laboratory, American Red Cross, 15601 Crabbs Branch Way,
Rockville, MD 20855, USA
FEATURES
Location/Qualifiers
1..1589
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="kidney"
1..1589
/gene="HCO001"
93..1478
/gene="HCO001"
/codon_start=1
/product="coronin-like protein"
/protein_id="AAA77058.1"
/db_xref="GI:1002923"
/translation="MSRQVRSKPRHVFQGPAPAKADQCYEDVRVSQTTWDSGFCVAVP
KEVALICEASGGAGFLVPLGKGTVDKNAPTVCGHTAPVLDIAWCHNDNVIASGSE
DCTVMWEIPDGLMLPREPVVTEGHTKRGIVAMHTTAQNVLSSAGCDNVLMVWD
VTGAAMTLGPEVHPDTIYSVDWSRDGLICTSCRDKRVRIIEPRKGTVAEKDRPH
EGTRPVRAVFSVEGKLTITGFSRMSERQVALWDTKHEEPLSLQELDTSSGVLLPFD
PDNTNYLCGGKDSIRYFEITSEAPFLHYSMFSSKESQRMGMXPXRGVLEVNKEI
ARFYKHERCEPIAMTVPRKSDLFQEDLYPTTAPGDPALTAEEWLGDRDAGPLLISL
KDGVPKPSRELKRVNRGLDTGRRRAAPEASGTPSSDAVSRLLEEEMRKQLQATVQELQK
LDRLKETVQAK"
300 a 492 c 494 g 277 t
BASE COUNT
ORIGIN
Alignment Scores:

```



```

/cell_type="lymphocyte-like cells"
1. .1639
/gene="CORO1A"
/notes="synonym: TACO"
123. 1508
/gene="CORO1A"
/notes="coronin-like protein p57; coronin 1A"
/codon_start=1
/product="tryptophane aspartate-containing coat protein"
/protein_id="AA18516.1"
/db_xref="GI:20271119"
/translation="MSRQVRSKFRHFVGPAPAKADQCYEDVVRVTSQTTWDSGFCVAVNP
KVALICASGGAGFLVPLGTVGRVDKNAPTVCGHTAPVLDIAMCPHNDNVIASGSE
DCTVMWEPDGLGLPLREPVVTLEGHTKRVGIAMHTTAQNVLISAGCDNMVIMVDP
VGTGAMLTGFEVPHDIYISVDMSDGLICTSCRDKVR.IIEPKRGIVVAEKORPH
ESTPRVAVVSEGLKITTFGRMSERQVALWDTKHLEPLSLQELDTSSGVLPPFPF
PTNIVYCGKDSIRIETSEAPFLHYLSMFSSKESQRMGMYPKGLGVNKKCEI
ARFYKHLHRCRFLANVTPRKSDQFEDLYPPTAGDPDALTAEWLGGRDAGPLLTIS
KDGYPVPKSRLEVRNGLDTGRRRAAPEASGTPSSDAVSRLEEMRKLTQATVQELQKR
LDRLEETVQAK"
BASE COUNT 336 a 502 c 503 g 298 t
ORIGIN

Alignment Scores:
Pred. No.: 104 Length: 1639
Score: 90.00 Matches: 40
Percent Similarity: 41.54% Conservative: 14
Best Local Similarity: 30.77% Mismatches: 41
Query Match: 12.71% Indels: 35
DB: Gaps: 6

US-10-087-573-2 (1-141) x AF495470 (1-1639)
QY 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26
Db 523 ACACGGCTTGCTGGGCCCTCCAGGGTGACACGGGTTCGCCGAGGGGCGACATCAGGCC 464
QY 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaValMetLeuArg 42
Db 463 CCATCCGGGATTTCCACACCATGACTGTGCGAGTCTCTCGAGCCCATGCGCAATGACGTTG 404
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
Db 403 TCATTG-----TGCGGGACCACGGCGATGTCTAGCACAGGGGCTGTGTGGCCACAG 353
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
Db 352 ACGGTGGCGCATTTCTTGTCACACGTCACAGTCCAGTCTTGCCCGAGGGGACACAGGAAGGCC 293
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db 292 CCTCCGCCG-----LysThrIleSerGlnGlu-SerAlaArgVa 108
QY 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108
Db 268 AGGGCCACAACTTAGGTTGACACAGAAAGCCACTGCTCCAGTGGTCTGTGAGACG 209
QY 108 LasnHisArgLeuProGluGlyHisPro 117
Db 208 CGCACATCTTCATAGCACTGGTCGGCCT 181

RESULT 24
E11332/c
LOCUS
DEFINITION
Human cDNA encoding a novel protein which has WD40 repeat and
leucine zipper.
E11332
ACCESSION
E11332.1 GI:22024974
VERSION
JP 1996119996-A/1.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

DEFINITION		Sequence 667 from Patent WO0164886.	
ACCESSION	AX237691		
VERSION	AX237691.1	GI:15797244	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 Gaiger A., Algate, P.A. and Mannion, J.		
AUTHORS	Compositions and methods for the detection, diagnosis and therapy of hematological malignancies		
TITLE	Patent: WO 0164886-A 667 07-SEP-2001;		
JOURNAL	CORIXA CORPORATION (US)		
FEATURES	Location/Qualifiers		
Source	1..1659		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	360 a	497 c	511 g 291 t
ORIGIN			
Alignment Scores:			
Pred. No.:	106	Length:	1659
Score:	90.00	Matches:	40
Percent Similarity:	41.54%	Conservative:	14
Best Local Similarity:	30.77%	Mismatches:	41
Query Match:	12.71%	Indels:	35
DB:	6	Gaps:	6
US-10-087-573-2 (1-141) x AX237691 (1-1659)			
QY	8	ThrAsnPreValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg	---
Db	501	ACACGCTGTGTGGCCCTCCAGGGTACGACGGGCTCCCGCAGGGGCACGATCAGGCC	442
QY	27	-----GluAlaLeuLeuArgValIysSerSerGluArgLeuAlaMetLeuArg	42
Db	441	CCATCGGGATCTCCACACCATGACTGTGTCAGTCTCCGGAGCCACTGGCAATGACGTTG	392
QY	43	AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla	---
Db	381	TCATTG-----TGGCGGCACACCGCATGCTAGCACGGGGCTGTGGCCACAG	331
QY	59	-----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu	74
Db	330	ACCGTGGCGCATCTTGTCCACACGTCCTGCTTGGCCAGGGGCAGCAGGAGGCC	271
QY	75	LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle	94
Db	270	CTCTCCCGCG-----CTGGCCTCACGATC	247
QY	95	ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa	108
Db	246	AGGGCCACAACTTAGGTTGACAGCAGACAGCCACTGTCCAGGTGTCTGTGAGACG	187
QY	108	IasnHisArgLeuProGluGlyHisPro	117
Db	186	CGCACATCTTCATAGCACTGGTGGCGCT	159
RESULT 26			
LOCUS	HUMPS7B/c		
DEFINITION	Human mRNA for actin binding protein p57, complete cds.		
ACCESSION	D44497		
VERSION	D44497.1		
KEYWORDS	p57; actin binding protein; WD repeat; leucine zipper.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 1659)		
	Suzuki, K., Nishihata, J., Arai, Y., Homma, N., Yamamoto, K., Irimura, T.		

and Toyoshima, S.  
Molecular cloning of a novel actin-binding protein, p57, with a WD repeat and a leucine zipper motif  
FEBS Lett. 364 (3), 283-288 (1995)  
95278344  
PUBMED 7758564  
2 (bases 1 to 1659)  
Suzuki, K.  
Direct Submission  
Submitted (26-DEC-1994) Kensuke Suzuki, Japan Tobacco Inc., Pharmaceutical Basic Research Lab.; 13-2, Fukuura 1-chome, Kanazawa-ku, Yokohama, Kanagawa 236, Japan  
(tel:045-786-7693(ex:3290), Fax:045-786-7692)  
Location/Qualifiers  
1. 1659  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/sub\_species="Caucasian"  
/db\_xref="taxon:9606"  
/sex="female"  
/cell\_line="HL60 (promyelocytic leukemia cell)"  
/tissue\_type="peripheral blood leukocytes"  
/clone\_lib="lambda gt10"  
/dev\_stage="36-years-old"  
101..1486  
CDS  
/codon\_start=1  
/product="human p57"  
/protein\_id="BA007940.1"  
/db\_xref="GI:927649"  
/translation="MSRQVRSKFRHFVQPAKQOCVEDRVSTTTWDSQFCAVNP  
KFTALICEAGSGGAFILPQKTVGDKNAPTCVGHATPVLDIANGPHNDNVIASGE  
DFTVMVEIPDGGAMFLPPEPVVTLEGTNRVGIWVHTTAQNVLLSAGCDNVIMWSD  
VGTGAAMLTVGPVHPDTIYSDVSRDGGGLICTSCRDKRVRIIEPRKGTWAEKDRPH  
EGTRPVRAVPSSEGLITTCFGRMSRQVALWDTKLEBPLSLQELDTSGVLLPFPD  
PDNIVYLCKGSSIRYFRITSEAPPLHVLMSFESSKESQSGRGVMPKRGLEVNKEI  
ARYKHLHERCEPIAMTVPKSLDFPQEDLPPTAGDPDPAALAEENLGGDAGPLLSL  
KDGVPKPSRVRNRLDTRRRRAPEASGTFSSDAVSKREEMKRLQATVQELQKR  
LDLEETVQAK"

BASE COUNT 360 a 497 c 511 g 291 t  
ORIGIN

Alignment Scores:  
Pred. No.: 106 Length: 1659  
Score: 90.00 Matches: 40  
Percent Similarity: 41.54% Conservative: 14  
Best Local Similarity: 30.77% Mismatches: 41  
Query Match: 12.71% Indels: 35  
DB: 9 Gaps: 6

US-10-087-573-2 (1-141) x HUMPS7B (1-1659)

Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26  
|||  
Db 501 ACACGCTTGGTGTGGCCCTCCAGGGTGACGACGGGCTCCCGCAGGGGCACATCAGGCC 442  
Qy 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
|||  
Db 441 CCATCCGGGATCTCCACACATGACTGTGCAGTCTCTGGAGGCATGGCAATGAGTTG 382  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
|||  
Db 381 TCATTG-----TGGGGGCACACGGCGATGTAGCAGGGGCTGTGTGGCCACAG 331  
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
|||  
Db 330 ACCGTGGGGCATCTTGTTCACACGTCAGTCTTGCACAGGGGCACACAGGAGGCC 271  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
|||  
Db 270 CCTCCCCCG-----CTGGCCTCACAGATC 247  
Qy 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108  
|||

Db 246 AGGCCACAACTTAGGTTGACACACAGAGCCACTGTCCACGGTGTGTGTGAGACG 187  
Qy 108 lAsnHisArgLeuProGluGlyHisPro 117  
|||  
Db 186 CCACATCTTATAGCACTGTGTGGCCT 159

RESULT 27  
AC120207 208443 bp DNA linear HTG 30-MAY-2002  
AC120207/c  
LOCUS  
DEFINITION  
Canis familiaris clone RP81-323G5, WORKING DRAFT SEQUENCE, 9  
ordered pieces.  
AC120207  
AC120207.2 GI:21263256  
HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Canis familiaris  
Canis familiaris  
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
REFERENCE  
1 (bases 1 to 208443)  
Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C.,  
Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J.,  
Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Laric, P.,  
Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,  
Margulies, E.H., Masiello, C., Maskeri, B., Mastrian, S.D.,  
McGloskey, J.C., McDowell, J., Paguirigan, C., Pearson, R.,  
Portnoy, M.E., Prasad, A., Schueler, M.G., Stantrop, S., Thomas, J.W.,  
Thomas, P.J., Touchman, J.W., Tsurguev, C., Vogt, J.L., Walker, M.A.,  
Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.  
NISC Comparative Sequencing Initiative  
Unpublished  
2 (bases 1 to 208443)  
Green, E.D.  
Direct Submission  
Submitted (04-MAY-2002) NIH Intramural Sequencing Center, 8717  
Grovmont Circle, Gaithersburg, MD 20877, USA  
REFERENCE  
3 (bases 1 to 208443)  
Green, E.D.  
Direct Submission  
Submitted (30-MAY-2002) NIH Intramural Sequencing Center, 8717  
Grovmont Circle, Gaithersburg, MD 20877, USA  
On May 30, 2002 this sequence version replaced gi:20451078.  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc.zoo@nhgri.nih.gov  
----- Project Information  
Center project name: CWO  
Center clone name: 323G05

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.  
----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 207000 bases at least Q40  
Consensus quality: 207366 bases at least Q30  
Consensus quality: 207588 bases at least Q20  
Insert size: 180000; agarose-fp  
Insert size: 207643; sum-of-ctngs  
Quality coverage: 9.29x in Q20 bases; agarose-fp

Quality coverage: 8.50x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1 24665: contig of 24665 bp in length  
24666: contig of 24665 bp in length  
24765: gap of unknown length  
37707: contig of 12942 bp in length  
37807: gap of unknown length  
37808: contig of 4186 bp in length  
41993: gap of unknown length  
42093: gap of unknown length  
51700: contig of 9607 bp in length  
51800: gap of unknown length  
79099: contig of 27299 bp in length  
79199: gap of unknown length  
115446: contig of 36247 bp in length  
115447: gap of unknown length  
150425: contig of 34879 bp in length  
150525: gap of unknown length  
171584: contig of 21059 bp in length  
171585: gap of unknown length  
208443: contig of 36759 bp in length.

Location/Qualifiers

1..208443  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9615"  
/clone\_lib="RP81-323G5"  
/clone\_lib="RP81"  
1..24665  
/note="assembly\_fragment"  
clone end: T7  
vector\_side: left  
24766..37707  
/note="assembly\_fragment"  
37808..41993  
/note="assembly\_fragment"  
42094..51700  
/note="assembly\_fragment"  
51801..79099  
/note="assembly\_fragment"  
79200..115446  
/note="assembly\_fragment"  
115447..150425  
/note="assembly\_fragment"  
150526..171584  
/note="assembly\_fragment"  
171585..208443  
/note="assembly\_fragment"  
clone end: SP6  
vector\_side: right  
55095 a 46990 c 47491 g 58067 t 800 others

Alignment Scores:

Pred. No.:	3,136+04	Length:	208443
Score:	90.00	Matches:	40
Percent Similarity:	39.46%	Conservative:	18
Best Local Similarity:	27.21%	Mismatches:	49
Query Match:	12.71%	Indels:	40
DB:	2	Gaps:	7

US-10-087-573-2 (1-141) x AC120207 (1-208443)

2 GluSerThrSerThrThrThrAspPheValAlaGluAsnArgProThrPheGlyGluThr 21  
|||||  
176099 GAACAGACAAACACAAACAAAC-----AAAAATCTCTCTACCTTCGGG---GCG 176052

Qy 22 PheAspValMetArgGluAlaLeuLeuArgValLys-----SerSer 35  
Db 176051 CTTACACTCTCGAGAGGGTTGGCTGATTTCACAGATGGTAGAGCTGAGGTGACGCTCA 175992  
Qy 36 GluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGly 55  
Db 175991 GAGAGGCTAAGCAACCTTGACCAAGGTCACGGGCGCTGCGGTTAAAGAATCAGGATTGGG 175932  
Qy 56 -----ThrGlyAlaSerAlaAlaAlaThrValThrProLys----- 68  
Db 175931 ACACAGGACTGTCTGACTCCAGATATGAGGAATGCAATTTCTACCAACACACAGATTACCT 175872  
Qy 69 -----GlyAlaSerMetLysLeuLysProArgProGln----- 80  
Db 175871 CTGGGATCGTGGGCTATCTGGAAGTGTCTCCCAACACCCCAATCTAACACAA 175812  
Qy 81 -----SerThrLysSerProGluLeuArgGluLeuSerArgLys 93  
Db 175811 TAAAGGAGACTTGGCAGAGATCCACACCCCGACTACCTTGGAACCTCTCTCCCTTGTGCG 175752  
Qy 94 IleArgGluMetAsnLysThrIle-----SerGlnGluSer 105  
Db 175751 GTATATGTGTTTAAACAAGATGTGAGAAATACCTGCCAGGTCTCTTCTTGGCCAGAGCCAA 175692  
Qy 106 AlaArgValAsnHisArgLeu 112  
Db 175691 CGACAGAGCCACACCGCAATT 175671

RESULT 28  
AB088224/c

LOCUS AB088224 210614 bp DNA linear BCT 11-JUN-2003  
DEFINITION Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.  
ACCESSION AB088224  
VERSION AB088224.1 GI:30698345

KEYWORDS Streptomyces rochei  
SOURCE Streptomyces rochei  
ORGANISM Streptomyces rochei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
AUTHORS Kinashi, H., Fujii, S., Hatani, A., Kurokawa, T. and Shinkawa, H.  
TITLE Physical mapping of the linear plasmid pSLA2-L and localization of the eryAI and actI homologs  
JOURNAL Biosci. Biotech. Biochem. 62, 1892-1897 (1998)  
AUTHORS Suwa, M., Sugino, H., Sasaki, A., Mori, E., Fujii, S., Shinkawa, H., Nimi, O. and Kinashi, H.  
TITLE Identification of two polyketide synthase gene clusters on the linear plasmid pSLA2-L in Streptomyces rochei  
JOURNAL Gene 246 (1-2), 123-131 (2000)  
MEDLINE 20231737  
PUBMED 10767533

REFERENCE 3  
AUTHORS Hiratsu, K., Mochizuki, S. and Kinashi, H.  
TITLE Cloning and analysis of the replication origin and the telomeres of the large linear plasmid pSLA2-L in Streptomyces rochei  
JOURNAL Mol. Gen. Genet. 263 (6), 1015-1021 (2000)  
MEDLINE 20408175  
PUBMED 10954087

REFERENCE 4  
AUTHORS Mochizuki, S., Hiratsu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K. and Kinashi, H.  
TITLE The large linear plasmid pSLA2-L of Streptomyces rochei has an unusually condensed gene organization for secondary metabolism  
JOURNAL Mol. Microbiol. 48 (6), 1501-1510 (2003)  
MEDLINE 12791134  
PUBMED 12791134

REFERENCE 5  
AUTHORS Mochizuki, S., Hiratsu, K. and Kinashi, H.  
TITLE Direct Submission  
JOURNAL Submitted (15-JUL-2002) Hanyasu Kinashi, Hiroshima University, Department of Molecular Biotechnology, Graduate School of Advanced

Sciences of Matter; 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8530, Japan [E-mail: kinashi@hiroshima-u.ac.jp, Tel: 81-824-24-7869, Fax: 81-824-24-7869]

The nucleotide sequence has been determined by using restriction fragments and nested deletion fragments of the ordered cosmid library of pSLA2-L.

pSLA2-L is 210,614 bp long with G-C content of 72.8 % and carries 143 ORFs.

Gene prediction was based on the unique codon usage in Streptomyces [Bibb et al., Gene 30:157-66 (1984)] using the FramePlot program of Ishikawa and Hotta (FEMS Microbiol Lett 174: 251-253 (1999)) as implemented at <http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl>. Where possible we chose an initiation codon (atg, gtg, ctg or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13 bp before the initiation codon). If this could not be identified we chose the most upstream initiation codon.

Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes (three PKS gene clusters for lankacidin, lankamycin and an unknown type II polyketide, a carotenoid biosynthetic gene cluster, and many regulatory genes and others). The range of each biosynthetic gene cluster has been deduced by comparison with similar gene clusters in most cases and may be revised in future.

Location/Qualifiers

```

1..210614
/organism="Streptomyces rochei"
/mol_type="genomic DNA"
/strain="7434AN4"
/db_xref="taxon:1928"
/plasmid="pSLA2-L"
/notes="linear plasmid"
1..1992
/notes="left terminal inverted repeat, TIR-L; shows 99.4 %
(1981/1992) sequence identity to TIR-R (complement
(208623..210614))"
683..2188
/notes="N-terminal sequence is almost identical (435/437)
with that of ORF143 at the right end of pSLA2-L until the
inner end of TIR-L"
ORF1 (501 aa)
similar to AL590463 Streptomyces coelicolor putative
helicase, SCPI_136 (879 aa); homology is seen until the
inner end of TIR-L"
/codon_start=1
/transl_table=11
/product="putative helicase"
/protein_id="BAC76459.1"
/db_xref="GI:30698346"
/translation="MSTTQRTQREAAQAEAVDVRALELPVRLAPERGLRTQVIM
ATSGTRVAARSAEKLRGRVLLVPSLDLLTQTEARWEAGRTGPMIGVSSLRGED
VAPPNTDVEELVDVWPDPKVTATYASLGTLERAHRCGLPCWDLIVVDEAHT
SGELKPWAVHDTNIPSLRLYNTATPRLQCLDEAGAGELVASMEDDPOGLFG
ARCTLTSEADRGCAPYQVVCVDITDTQLQAQLLQVEGRSDVRGAALQTA
LKASSEENRFTLVFHHVKEAEAPAGLPVAKRLHAAGFLYPTIWNALCGEH
KPHRRVLGFTSGIATDGTVEKGLSVKLVGEGVDRECDSVIADVRGSMPLD
VOAVGRLRIQPCQGVASLVVPLVLLDPCGTADNMLTSRPYGLALLFRQPLLTGHG
EEGFRAARTPGFEVRPLGVGRGAQSGASAGPASRTTSPMSLSISMRA"
complement(2315..3595)
/notes="ORF2 (426 aa)
similar to AE004736-10 Pseudomonas aeruginosa hypothetical
protein (442 aa)"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="BAC76460.1"
/db_xref="GI:30698347"
/translation="MRKRRALLGLTGLLLVAVLAPAPAAHASADTYVGTWAAPTAA
PADSTVYEDQLQRTVHLSVAGDSLRVFTNEFTGSLTIGEVAHAPAAQGPATAV
DPTDRVFGKRPSTALPAGTQVRSDPVALPTAGGDLVLSLYLPQTGTSVHSA
YGHFVAAGDVTGAPDLTSTATSHFSLGVSVDVRAGTADSVVITLGDSTIDGHT
TLDNRWDLALAEIRLRDGLAGTGVNAG:GGNRLLRDPDPFGSAESGFAAYFGE
SALKRFDVLPQGARAVTVLLGVNDLQPGIAAPASDEVTAEELIAGYQLIERAH
EHLKIYGATITFPAGDTI:GYFTPREAVRQVNDWIRTSAGFDIVLDFDAVLDPAR

```

PDHLLPAYDGGDLHPNDAGMAAARAFPLDLSLR"

complement(3659..4645)

/notes="ORF3 (328 aa)

similar to Y00459-2 Streptomyces griseus regulatory protein, StrR (350 aa)"

/codon\_start=1

/transl\_table=11

/product="strr-like regulatory protein"

/protein\_id="BAC76461.1"

/db\_xref="GI:30698348"

/translation="MTAALAGNLDQETRSIPSSLPVGYSPRLCGEADAEHVARLDPD

EALPQPERAQRILIAQPHLSHRAVAGIAGTAAALRPAEAAGFTLVRSNMTHGL

PLQSPERARQIRLVGDHMRVLAAKAGHTTIEVRLFDGAAEAGFTLVRSNMTHGL

GRIRPLDCTGGRKAAELLAQRQASVREVARHAGISPATASVDRRLASGRSPVPR

ITPGARPAGTGGRAIPGGGAPEPVSVPFIRPPDSPLVRLLRDPSLSRSGRKL

LRLLQCCAVERTALLAAQTVPFCHTDLVLAELAREYADLWAEFAREVAR"DG"

complement(4922..5812)

/notes="ORF4 (296 aa), lankacidin biosynthesis protein

similar to AE004625-4 Pseudomonas aeruginosa

pyrroloquinoline quinone biosynthesis protein B (304 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein B"

/protein\_id="BAC76462.1"

/db\_xref="GI:30698349"

/translation="MILLTGAAGGPPQWNCACALCARRRGELPARSQECVAVSGDG

RDWLNASPDITQLLAALTEGPGPRTDTPVRGVLITDAEVDHALGLAVRGATGL

TVYAAPVPGALSAELPVRGLLORYAPWMDRDTAPGFAVAGLTVTAHPVGTAKPK

YAHAPDPAPWVCARIEDPATGALVYAFCLATWPDGPDLLASATCALLDGTFESA

GELGTSSAGAGQSLMGHLPVAGPGGSLAALARHRLRIYTHLNNTELLDPPSSAA

HAAVREAGVEVLPDGSSELV"

complement(5815..6909)

/notes="ORF5 (364 aa), lankacidin biosynthesis protein

similar to AJ277117-5 Gluconobacter oxydans putative pqqE

protein (359 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein B"

/protein\_id="BAC76463.1"

/db\_xref="GI:30698350"

/translation="MADPAVAGAPAGLIELTRCPHCPVCSNPLELVREAEITCQO

XTDILTQARELGVOVQHFSGEELARPDLPDLVGHARRLGAYVNLVTSVGLTAERAH

DLARGVDVQLSLQDAPAAQAIAGARVHTAKLEAARAVTATAGLPLTVILHRRN

IDRTGRMVDLADVLGDRIELANTQYVGLGRNRAALMPTAAQAAAREAVHARTY

AGGELYVAAVDYDRPKFCMDGWSGTQVTPAGDVLPCPAAYITLTPVENALRR

PLSITVYASRSFNAYRGTMWRPRTCTCPERHADHGGRCQAFQLTGDAATDPACGL

SPHRSVDAALAEVTDGVPVAFVPRGPVPA"

complement(6902..7174)

/notes="ORF6 (90 aa), lankacidin biosynthesis protein

similar to AL603642-200 Sinorhizobium meliloti putative

pyrroloquinoline quinone biosynthesis protein D (98 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein D"

/protein\_id="BAC76464.1"

/db\_xref="GI:30698351"

/translation="MTGLPEPTVPRLPVRLTRDPARGSLALLPERVVVINDTAAAV

LARCDGTSIAGIVERLAEYEGVSAEDVRELLRLAQRVVDLHG"

complement(7171..7899)

/notes="ORF7 (242 aa), lankacidin biosynthesis protein

similar to AE004625-5 Pseudomonas aeruginosa

pyrroloquinoline quinone biosynthesis protein C (250 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein C"

/protein\_id="BAC76465.1"

/db\_xref="GI:30698352"

/translation="MSMSTVREAAAPMSEAEFRQLHALSSSYMDRPFHRRMHEGILL

DEGELRLWAANRYQRCLEPKQDAAIWANGPLPEVRROWLSRIYVHDGACAGAGAEK

WRLRLAEAVGRROEDERLVLACTREAVDVFARRRPWLAAASGLTELSPGLL

AHLGRLEHRYPIAESEGFYFETAEVVGFEGRSLDLVLAARVAREQEQEACVRALA

FKKRVLNAVDSLDIHTGNGATRS"

complement(7926..8030)

COMMENT

LTR

CDS

CDS

CDS

```

/note="ORF8 (34 aa), lankacidin biosynthesis protein
similar to AL603642-197 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein A (31 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein A"
/protein_id="BAC78466.1"
/db_xref="GI:30698353"
/translation="MRTSGKELPAKAWHRPDEFVTIDTGMETAYFSR"
8420..10057
CDS
/note="ORF9 (545 aa), possible lankacidin resistance
protein"

```

```

Alignment Scores:
Pred. No.: 3.17e+04 Length: 210614
Score: 90.00 Matches: 41
Percent Similarity: 39.74% Conservative: 19
Best Local Similarity: 27.15% Mismatches: 57
Query Match: 12.71% Indels: 34
DB: 1 Gaps: 5

```

US-10-087-573-2 (1-141) x AB088224 (1-210614)

```

QY 5 SerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspVal 24
Db 190198 TCACCCGCGCGCGCAGCCCGATCGATAGCCCAACGACGCTACTGCGCGACGCT 190139
QY 25 MetArgGluAlaLeu----- 29
Db 190138 CTAAGAGAGGCGCTTCACGCGAGTGGCGCGCGTATCGCGCGACCCACTCACT 190079
QY 30 -----LeuArgValIysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMet 47
Db 190078 CTCGCGTGGCGGCTCAACGCGACCTGACGCTCTATCGGCTACCGCGGTACCGCGGCGCT 190019
QY 48 CysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrPro 67
Db 190018 TCGGCCACCGC---TGTCGCGAGTGGCGAGCAGCAAGCCCGACGACCAACAGCA 189962
QY 68 LysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSerProGluLeu 87
Db 189961 CGATCA-----CCGCGATCCACACAGCGCGCTTCACGG 189929
QY 89 ArgGluLeuSerArg--LysIleArgGluMetAsnLysThrIleSerGlnGluSerAla 107
Db 189928 CGAAACCCCGACCGAGAGATCAGACCCAGCAGAGAACCAAGACGAGGAGACCATCG 189869
QY 107 IgValAsnHisArgLeuProGluGlyHisProLeuLeuGlu-----LysArgA 123
Db 189868 TTATCAAC---CTCTGGCCACCGGATGCCCTTACTCAAGCCGACAGCCGCGGAACG 189812
QY 123 laGluTyrPheArgHisLeuArgSerLeu 132
Db 189811 AGAGTTATTTCGCGCGGTCCGGGGCATA 189783

```

```

RESULT 29
LOCUS AC012363/c 217346 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-438012 from 2, complete sequence.
ACCESSION AC012363
VERSION AC012363.6 GI:15778776
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

```

REFERENCE
1 (bases 1 to 217346)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 217346)

```

```

AUTHORS Ali, J., Kozlowski, A. and Dignan, G.
TITLE The sequence of Homo sapiens BAC clone RP11-438012
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 217346)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 217346)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 217346)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Sep 26, 2001 this sequence version replaced gi:15144350.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0438012
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-347L11. Actual start of this clone is at base position 1 of RP11-438012; actual end is at base position 217346 of RP11-438012.

Data from AC079879 was used to finish this clone, AC012363. Polymorphisms exist between AC012363 and AC079879. There is a single plasmid region from 188387 to 188395.

#### FEATURES

```

source
1..217346
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"

```

```

/repeat_region /map="2"
/repeat_region /clone="rp11-438012"
/repeat_region /clone_lib="RPCI-11"
/repeat_region 1..525
/repeat_region /rpt_family="L1"
/repeat_region 545..538
/repeat_region /rpt_family="AT_rich"
/repeat_region 548..1421
/repeat_region /rpt_family="L1"
/repeat_region 212..2522
/repeat_region /rpt_family="Alu"
/repeat_region 3888..4168
/repeat_region /rpt_family="Alu"
/repeat_region 4149..4168
/repeat_region /rpt_family="(A)n"
/repeat_region 4566..4950
/repeat_region /rpt_family="L1"
/repeat_region 4951..5079
/repeat_region /rpt_family="L1"
/repeat_region 5132..5499
/repeat_region /rpt_family="L1"
/repeat_region 5821..6117
/repeat_region /rpt_family="Alu"
/repeat_region 6368..6706
/repeat_region /rpt_family="MER1_type"
/misc_feature 6443..6447
/misc_feature /notes="match to EST T31897 (NID:g613995)"
/repeat_region 6750..6834
/repeat_region /rpt_family="Alu"
/repeat_region 6837..6893
/repeat_region /rpt_family="L2"
/misc_feature 7176..7744
/misc_feature /notes="match to EST AWB52630 (NID:g7948147)"
/repeat_region 7378..7505
/repeat_region /rpt_family="L2"
/repeat_region 7699..7977
/repeat_region /rpt_family="Alu"
/misc_feature 8308..8336
/misc_feature /notes="match to EST BF752822 (NID:gl2079498)"
/repeat_region 8398..8680
/repeat_region /rpt_family="Alu"
/repeat_region 8682..8725
/repeat_region /rpt_family="(T)n"
/repeat_region 8691..8921
/repeat_region /rpt_family="Alu"
/repeat_region 8935..9235
/repeat_region /rpt_family="Alu"
/misc_feature 8964..9480
/misc_feature /notes="match to EST BI085105 (NID:gl4503435)"
/misc_feature 9103..9598
/misc_feature /notes="match to EST BI089781 (NID:gl4508111)"
/misc_feature 9735..10199
/misc_feature /notes="match to EST AI684181 (NID:g4895475) tw96d01.x1"
/repeat_region 9921..9949
/repeat_region /rpt_family="AT_rich"
/repeat_region 10024..10362
/repeat_region /rpt_family="MER2_type"
/repeat_region 10412..10719
/repeat_region /rpt_family="Alu"
/repeat_region 10693..10719
/repeat_region /rpt_family="(A)n"
/repeat_region 10831..11014
/repeat_region /rpt_family="MER1_type"
/repeat_region 11186..11459
/repeat_region /rpt_family="Alu"
/misc_feature 11250..11276
/misc_feature /notes="similar to Homo sapiens EST BF769297 (NID:gl2117197)"
/repeat_region 13494..13698
/repeat_region /rpt_family="L2"
/repeat_region 13855..14300
/repeat_region /rpt_family="L1"
/repeat_region 14303..14359

/rpt_family="CT-rich"
14385..14949
/rpt_family="L1"
14950..15506
/rpt_family="MaLR"
15556..15731
/rpt_family="L1"
15738..15788
/rpt_family="MER1_type"
15819..17702
/rpt_family="L1"
17706..17806
/rpt_family="Alu"
17831..18050
/rpt_family="L1"
18094..18211
/rpt_family="MER1_type"
18242..18398
/rpt_family="L1"
18418..18702
/rpt_family="Alu"
18899..19204
/rpt_family="Alu"
19634..19891
/notes="match to EST AI824843 (NID:g5445514) wb02e10.x1"
20004..20582
/notes="match to EST AI769494 (NID:g5236003) wh57f08.x1"
20414..20786
/notes="match to EST AL043595 (NID:g5422982)"
20747..20748
/notes="match to EST U46342 (NID:gl236484)"
21412..21745
/notes="match to EST U46342 (NID:gl236484)"
22152..22728
/notes="match to EST AL042860 (NID:g5935608)"
22164..22441
/notes="match to EST C18353 (NID:gl579955)"
22359..22868

Alignment Scores:
Pred. No.: 3.29e+04 Length: 217346
Score: 90.00 Matches: 36
Percent Similarity: 38.46% Conservative: 9
Best Local Similarity: 30.77% Mismatches: 56
Query Match: 12.71% Indels: 16
DB: 9 Gaps: 4

US-10-087-573-2 (1-141) x AC012363 (1-217346)
QY 16 ProThrPheGlyGluThrPheAspValMetArgGluAlaLeuLeuArgValIysSerSer 35
Db 97847 CCGGCTCGGGCGGACATGCTCTGCGCGCGCGCCGACAGCCGCGGAGGAGC 97788
QY 36 GluArgLeu-----AlaMetLeuArgAlaLeuAlaGly 46
Db 97787 AAGAGCTCGGGCGGCGGTACCCACATCCCGGCGCGCGCGCGCGCGCGCGG 97728
QY 47 MetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThr 66
Db 97727 CCGCGGGCGCGCGCGCTTACGAGGACAGTCCCGCGCTCGCGCTCGGACCCGG 97668
QY 67 ProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerProGlu 86
Db 97667 CCC-----GCAGCTCTAAACGCGCGCGCGCGCTCCCGGCGCGCGCGCGCG 97614
QY 87 LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAla 106
Db 97613 TCCACCGATTGTTCATTCAGCTGCGGCTGCGGCTGCTTCCACC-----CCCCGAGCG 97563
QY 107 ArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAla 123
Db 97562 CGCTCGCGCGTCGCCTA-----GGCAACCCCTCTCTCGAGAGCGCGCC 97518

```

Qy	59	SerAlaIleAlaIaThrValThrProLysGlyAlaSerMetLysLeuLysProProArg	78
Db	756	GAAGCCGGAAGCCCGCGATCCCCCGATGGCGGCCAGCAGGCGTCCCGCAGATCAGCGC	697
Qy	79	ProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn	98
Db	696	CCCCGGCATGCCACGCCCATGGTGGCAAGAGTGGCAGGCAGTCCAGCACC GCCCG	637
Qy	99	LysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeu	118
Db	636	CGCCCCCTTCCCGTCCGGCGCAGCAACGAGATCGACGAATAGCCGCGAGCGCAGCAGGT	577
Qy	119	LeuGluLysArgAla-----GluTyrPheArgHisLeuArgSerLeu	132
Db	576	CGACAGCAGCCGCCGCCCAAGGGGCCAGCCGCCAGCAGCGGCACATGCCGCCGGTGC	517
Qy	133	LysSerGlnGly	136
Db	516	CATGGCCAGGGC	505
RESULT	31		
AC121492/c			
LOCUS			
DEFINITION	AC121492	96236 bp	DNA linear INV 20-MAY-2003
			Leishmania major chromosome 27 clone LB03590 strain Friedlin,
			complete sequence.
ACCESSION	AC121492		
VERSION	AC121492.3	GI:30911132	
KEYWORDS	HTG.		
SOURCE			
ORGANISM			
			Leishmania major
			Leishmania major
			Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
			Leishmania.
REFERENCE	1	(bases 1 to 96236)	
AUTHORS	Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,		
	McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-MAY-2002) Seattle Biomedical Research Institution, 4		
	Nickerson Street, Seattle, WA 98109-1651, USA		
REFERENCE	2	(bases 1 to 96236)	
AUTHORS	Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,		
	McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.		
	Direct Submission		
TITLE	Submitted (22-NOV-2002) Seattle Biomedical Research Institution, 4		
JOURNAL	Nickerson Street, Seattle, WA 98109-1651, USA		
REFERENCE	3	(bases 1 to 96236)	
AUTHORS	Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,		
	McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.		
	Direct Submission		
TITLE	Submitted (20-MAY-2003) Seattle Biomedical Research Institution, 4		
JOURNAL	Nickerson Street, Seattle, WA 98109-1651, USA		
COMMENT	On May 20, 2003 this sequence version replaced gi:25167130.		



Qy 4 ThrSerThrThrThrAsnPhaValAlaGluAsnArgProThrPheGlyGluThrPheasp 23  
Db 83917 TCCTCTTCATCACCAACTCCGCCAGCTGGCAGCGGCTCTTC---AGCATGTTCTCC 83861  
Qy 24 ValMetArgGluAlaLeuLeuArgVallySerSerGluArgLeuAlaMeLeuArgAla 43  
Db 83860 GCCTTCCTACAGCAGCAGCAGCTGTACCGGGTCAAGTAACCTTCATCATCTCGCTCT 83801  
Qy 44 Leu-----AlaGlyMetCysGlyHisArgValleu 53  
Db 83800 ACCTCCAGCGGTGGAACACATCGTCTTCATCAGCGGAATATGTCAGCAGATATTG 83741  
Qy 54 ProGlyThrGlyAlaSerAlaIleAlaIleAlaIle-----ValThrPro 67  
Db 83740 CCGCGCTGGGTGAGGTAGCTCGATCGCTCGATCGCTCGCAGCAACCTCGCGCCACCTCG 83681  
Qy 68 LysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSerProGlu--- 86  
Db 83680 AAGCGCGGTCTATTCAAGCCGATCCCTTCTCCGAGAAACAACAGCTTCCCAAGGA 83621  
Qy 87 ----LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSer 105  
Db 83620 TTCTTCGGAAGTTTCGCAAGTCAATTGATGAGCGCGCAGCAACACTGTCTGTGCAAA 83561  
Qy 106 AlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyr 125  
Db 83560 TCAGCACTCTCTGCAAGACCAAGGAGGAGCCCTCAACAGAGCAAGGTCCTCA--- 83504  
Qy 126 PheArgHisLeuArg-----SerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 83503 ACGCGAAGGCTACGCGCAACTCCGCTGCTCTGCGTTTAGCGCAGCATCCAGGCTG 83447  
RESULT 32  
AC084864/C 176562 bp DNA linear PRI 26-MAY-2002  
LOCUS Homo sapiens chromosome 7 clone RP11-738B7, complete sequence.  
AC084864  
ACCESSION AC084864  
VERSION AC084864.4 GI:21217402  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 176562)  
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.  
and Haugen, E.D.  
Direct Submission  
Unpublished  
2 (bases 1 to 176562)  
Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J. and Haugen, E.  
Direct Submission  
Submitted (23-NOV-2000) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
REFERENCE 3 (bases 1 to 176562)  
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C.,  
Clendenning, J., Ivey, R.G. and Haugen, E.D.  
Direct Submission  
Submitted (19-APR-2001) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
REFERENCE 4 (bases 1 to 176562)  
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.  
Direct Submission  
Submitted (06-APR-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
REFERENCE 5 (bases 1 to 176562)  
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.  
and Haugen, E.D.  
Direct Submission  
Submitted (26-MAY-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA

On May 26, 2002 this sequence version replaced gi:20066284.  
----- Genome Center  
Center: University of Washington Genome Center  
Center Code: UWGC  
Web site: http://www.genome.washington.edu  
Contact: uwgctg@u.washington.edu  
----- Project Information  
----- Project name: chr-7  
Center clone name: RP11-738B7 (djs722)  
----- Summary Statistics  
Assembly program: Phrap; version 0.990319  
Consensus quality: 176414 bases at least Q40  
Consensus quality: 176547 bases at least Q30  
Consensus quality: 176562 bases at least Q20  
Insert size: 176562; sum-of-contigs  
Quality coverage: 12.0x in Q20 bases; sum-of-contigs  
-----  
Overlapping Sequences:  
5': RP11-48A19 (UWGC:djs705) AC078846 58194-bp overlap  
3': RP11-437L1 (UWGC:djs723) AC084865 98271-bp overlap  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
Genbank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
This sequence was finished as follows unless otherwise noted:  
all regions were either double-stranded or sequenced with an  
alternate chemistry or covered by high quality data (i.e., Phred  
quality >= 30); an attempt was made to resolve all sequencing  
problems, such as compressions and repeats; all regions were  
covered by at least one plasmid subclone or more than one M13  
subclone; and the assembly was confirmed by restriction digest.  
-----  
Sequence Validation:  
This sequence has been validated by Multiple Complete Digest  
fingerprinting. Comparison of the experimentally derived digest  
fragments with sequence-predicted fragments is given below.  
The electronically-digested sequence consists of both insert and  
vector, in order to accurately represent the entire circular BAC.  
Small fragments below a variable cutoff (approximately 400-800 bp)  
are not resolved in the fingerprint and hence do not appear  
in the table. There are no significant remaining discrepancies  
between the experimental and predicted values. Uniquely ordered  
fragments are separated by dashed lines.  
-----  
EcoRI HindIII BglII  
SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt  
-----  
8696 8837 9894 9869 5769 5639  
-----  
6 <800 6382 6431 2067 2056  
-----  
3072 3060 512 <800 6151 6222  
-----  
12973 12654 449 <800 135 <800  
-----  
1184 1186 22215 21921 17291 16959  
-----  
13704 13524 14723 14460 4160 4079  
-----  
4548 4514 732 772 6476 6481  
-----  
1422 1494 300 <800 89 <800  
-----  
10713 10620 261 <800 2525 2559  
-----

7612	7664	3459	3373	681	<800	183	<800	3024	3237
1478	1494	597	<800	1924	1902	Location/Qualifiers			
1482	1494	2272	2282	3456	3411	1. 176562			
7824	8087	1345	1347	1474	1451	/organism="Homo sapiens"			
1721	1744	2481	2480	2724	2728	/mol_type="genomic DNA"			
313	<800	2806	2809	12250	12362	/db_xref="taxon:9606"			
10245	10620	594	<800	8808	8991	/chromosome="7"			
2957	2902	1091	1115	3945	4079	/clone_lib="RP11-738B7"			
4167	4235	1129	1115	4649	4766	/standard_name="sWS53347"			
6118	6170	2198	2282	1854	1902	47363. 47572			
7425	7499	4177	4278	7735	7953	56730. 56808			
1918	2065	24114	24615	19	<800	/standard_name="sWS53322"			
3930	3930	3744	3797	920	915	121881. 122003			
24765	24846	2477	2480	1950	1902	Alignment Scores:			
10168	10620	1254	1273	7997	7953	Pred. No.: 3.48e+04			
543	<800	7662	7831	12518	12362	Score: 88.50			
587	<800	1397	1507	17895	18146	Percent Similarity: 36.92%			
562	<800	3724	3797	2611	2728	Best Local Similarity: 26.15%			
10971	10620	1272	1273	684	<800	Query Match: 12.50%			
4315	4235	3335	3373	2714	2728	DB: 9			
5736	5630	2806	2809	1157	1158	US-10-087-573-2 (1-141) x AC084864 (1-176562)			
748	764	310	<800	5200	5136	Qy 44 LeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIle----- 61			
2861	2902	777	772	1038	1041	Db 70445 CTGGAGGGGGCTCGCGAGAGCGCGGGCGCGAGCTCGCGCTCCCGCTTCTTCA 70386			
10495	10620	9726	9869	2986	2986	Qy 62 -----AlaAlaThrValThrProLysGlyAlaSerMetLysLeu 74			
7	<800	20519	20229	327	<800	Db 70385 GGGTGGCGGCTGGTTCGAGGCTGCTTCGTCGCGCGGAGCGCGCGCCAGTCGTG 70326			
		3821	3797	81	<800	Qy 75 LysProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94			
		22	<800	5265	5136	Db 70325 GGGAAAGCGGCTCTCAGACCTCCCGTGCCTCCACTTCGC-----GGCCACTCATCC 70272			
		454	<800	3434	3411	Qy 95 ArgGluMetAsnLysThrIleSerGlnLysAlaArgValAsnHisArgLeuPro--- 113			
		6414	6431	900	915	Db 70271 CGGGAACCCGCGCACG-----CGAATGAGACACCGTTCGCCGCC 70230			
		907	915	34	<800	Qy 113 ----- 113			
		2334	2480	470	<800	Db 70229 TCAACTGCCCATTCCTGTTAATAGCGCTAGTGGCACCCAGGTTCCAGCCCTTCCCCAA 70170			
		284	<800	3878	3862	Qy 114 -----GlutGlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHisLeu 129			
		3580	3509	3274	3411	Db 70169 GATCGCGGGCTGACGCGCTTCCTCTGGCGCTCACCAGTCA-----GGCATCCC 70119			
		150	<800	8043	7953	Qy 130 ArgSerLeuLysSerGlnGlyValAsnArg 139			
		1262	1273	1863	1902	Db 70118 AGGACTATGGCAGTGAGGTGTCAACCCG 70089			
		5121	5107	2821	2986	RESULT 33			
						AE001863			
						LOCUS			
						DEFINITION			
						Deinococcus radiodurans R1 section 2 of 2 of the complete			
						chromosome 2.			
						ACCESSION			
						AE001863 AE001825			
						VERSION			
						AE001863.1 GI:6460670			
						KEYWORDS			
						SOURCE			
						ORGANISM			
						Deinococcus radiodurans			
						Deinococcus radiodurans			
						Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;			
						Deinococcaceae; Deinococcus.			
						REFERENCE			
						1 (bases 1 to 198677)			
						White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,			
						Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,			
						Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,			
						Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zaleski, C.,			
						Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.			
						Genome sequence of the radioresistant bacterium Deinococcus			
						TITLE			

radiodurans R1  
Science 286 (5444), 1571-1577 (1999)  
20036896  
10567266  
2 (bases 1 to 198677)  
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Mofatt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Nelson, K.E., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Frazer, C.M.  
Direct Submission  
Submitted (02-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1. 198677  
/organism="Deinococcus radiodurans"  
/mol\_type="genomic DNA"  
/strain="R1"  
/db\_xref="taxon:1299"  
/chromosome="2"  
/complement (91..1926)  
/gene="DRA0206"  
/complement (91..1926)  
/gene="DRA0206"  
/feature="similar to GB:AE000520 percent identity: 43.92; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="oligodeoxyphosphatase F, putative"  
/protein\_id="AAFI2423.1"  
/db\_xref="GI:6460718"  
/translation="MLARACPLRGACPRSCRPAPOTSCPGWERERRWARLPMTTQ SNLSEVHKLVNPAADQASAPAEFAALQEAIDLTAQDVPAWLRNLSLBARLQSGS KLSTHADLTHDQAIQRYQTFVAEVLPAQERASQALTEKLAIVPGVPGPFPALNR RFQQAALFRFANVDLGVTHNQSHSVITGNQVGLQCELTLPQAKQGGSPDRA QREAAWALSADNAIAPELDVMDLIRTRQLAHNADENFRWKELEDRVDYTP EQCLDFVSRVDEVPVLLSGVLGDIARQLGLESRPWDYNNRNLDPGEGESLRFPT GADLEALQESLDLPDLAARFQMRSEGLLDLESRPGKWTAYCNYPTTNEPFL MNVGTAEDEVVLVEHGAHPGFGYSGDRQPLVNNRWSPIEFVEIPSMAEFTLDEL SHVSDDLARVRSKQLEGVITAFIPMAAQDAFQHWLYAEPEDVTIADLAKWLELD RTFHFFDQKSLGDSVRKAGWYVHIQVQPYEYEVAMCYLAARVAIVRAARQDPAAL GRYKDALGNVTGVPELYRAAGAEFFREYINGLMFLKEQLG"  
/complement (2005..3645)  
/gene="DRA0207"  
/complement (2005..3645)  
/gene="DRA0207"  
/feature="similar to PID:1653945 percent identity: 49.28; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI2422.1"  
/db\_xref="GI:6460717"  
/translation="MNFEEVLSFRLSSVAHIALNRASLKEEMSKRSLSALLGALL SACSPENGPQAPQLPFPTEAPGTPSGKRGVLWDVAFGPGIKTPEEVSTLUDTAR AMVNTLVQVGRGDCYCNNAAPRTDDPEVPAGDFADVIKAKAHADIEVNAII TADVNSVTTPKSAHEAFNTHGPNAGRDNLVTKADGTAKAGADWLDPGHDAAY TRNMYVSAKAYVDGIMEDRVRYPDYNEUGVQWQCYNETALRYAEATGATGDDP ADPQASNRQGVNLRVETALAKAYKPGVAIAAATITYGEPDAEAPFAASRPYAE VQCDMLTKRGYLDQNNVMYKRDFFVLQKQWFDQNTTFAAGLRQRYPIEVOVSGAA IYLNDDIASVQIRKTAAGLGWAGYRYTPDKDVGEGKRAAEVADLKLRLTDAGG PFPTTQKAPLFAASVRLGSELGELRAARPGHHSVELVRAGGTWVDVAATFSDQGRYG FLNLPQGAFFDVRVDGKILLPNVQATPGRTVTMLPITLN"  
/complement (3751..4860)  
/gene="DRA0208"  
/complement (3751..4860)  
/gene="DRA0208"  
/feature="similar to PID:1001216 PID:1001269 percent identity: 66.97; identified by sequence similarity; putative"  
/codon\_start=1

radiodurans R1  
Science 286 (5444), 1571-1577 (1999)  
20036896  
10567266  
2 (bases 1 to 198677)  
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Mofatt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Nelson, K.E., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Frazer, C.M.  
Direct Submission  
Submitted (02-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1. 198677  
/organism="Deinococcus radiodurans"  
/mol\_type="genomic DNA"  
/strain="R1"  
/db\_xref="taxon:1299"  
/chromosome="2"  
/complement (91..1926)  
/gene="DRA0206"  
/complement (91..1926)  
/gene="DRA0206"  
/feature="similar to GB:AE000520 percent identity: 43.92; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="oligodeoxyphosphatase F, putative"  
/protein\_id="AAFI2423.1"  
/db\_xref="GI:6460718"  
/translation="MLARACPLRGACPRSCRPAPOTSCPGWERERRWARLPMTTQ SNLSEVHKLVNPAADQASAPAEFAALQEAIDLTAQDVPAWLRNLSLBARLQSGS KLSTHADLTHDQAIQRYQTFVAEVLPAQERASQALTEKLAIVPGVPGPFPALNR RFQQAALFRFANVDLGVTHNQSHSVITGNQVGLQCELTLPQAKQGGSPDRA QREAAWALSADNAIAPELDVMDLIRTRQLAHNADENFRWKELEDRVDYTP EQCLDFVSRVDEVPVLLSGVLGDIARQLGLESRPWDYNNRNLDPGEGESLRFPT GADLEALQESLDLPDLAARFQMRSEGLLDLESRPGKWTAYCNYPTTNEPFL MNVGTAEDEVVLVEHGAHPGFGYSGDRQPLVNNRWSPIEFVEIPSMAEFTLDEL SHVSDDLARVRSKQLEGVITAFIPMAAQDAFQHWLYAEPEDVTIADLAKWLELD RTFHFFDQKSLGDSVRKAGWYVHIQVQPYEYEVAMCYLAARVAIVRAARQDPAAL GRYKDALGNVTGVPELYRAAGAEFFREYINGLMFLKEQLG"  
/complement (2005..3645)  
/gene="DRA0207"  
/complement (2005..3645)  
/gene="DRA0207"  
/feature="similar to PID:1653945 percent identity: 49.28; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI2422.1"  
/db\_xref="GI:6460717"  
/translation="MNFEEVLSFRLSSVAHIALNRASLKEEMSKRSLSALLGALL SACSPENGPQAPQLPFPTEAPGTPSGKRGVLWDVAFGPGIKTPEEVSTLUDTAR AMVNTLVQVGRGDCYCNNAAPRTDDPEVPAGDFADVIKAKAHADIEVNAII TADVNSVTTPKSAHEAFNTHGPNAGRDNLVTKADGTAKAGADWLDPGHDAAY TRNMYVSAKAYVDGIMEDRVRYPDYNEUGVQWQCYNETALRYAEATGATGDDP ADPQASNRQGVNLRVETALAKAYKPGVAIAAATITYGEPDAEAPFAASRPYAE VQCDMLTKRGYLDQNNVMYKRDFFVLQKQWFDQNTTFAAGLRQRYPIEVOVSGAA IYLNDDIASVQIRKTAAGLGWAGYRYTPDKDVGEGKRAAEVADLKLRLTDAGG PFPTTQKAPLFAASVRLGSELGELRAARPGHHSVELVRAGGTWVDVAATFSDQGRYG FLNLPQGAFFDVRVDGKILLPNVQATPGRTVTMLPITLN"  
/complement (3751..4860)  
/gene="DRA0208"  
/complement (3751..4860)  
/gene="DRA0208"  
/feature="similar to PID:1001216 PID:1001269 percent identity: 66.97; identified by sequence similarity; putative"  
/codon\_start=1

radiodurans R1  
Science 286 (5444), 1571-1577 (1999)  
20036896  
10567266  
2 (bases 1 to 198677)  
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Mofatt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Nelson, K.E., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Frazer, C.M.  
Direct Submission  
Submitted (02-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1. 198677  
/organism="Deinococcus radiodurans"  
/mol\_type="genomic DNA"  
/strain="R1"  
/db\_xref="taxon:1299"  
/chromosome="2"  
/complement (91..1926)  
/gene="DRA0206"  
/complement (91..1926)  
/gene="DRA0206"  
/feature="similar to GB:AE000520 percent identity: 43.92; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="oligodeoxyphosphatase F, putative"  
/protein\_id="AAFI2423.1"  
/db\_xref="GI:6460718"  
/translation="MLARACPLRGACPRSCRPAPOTSCPGWERERRWARLPMTTQ SNLSEVHKLVNPAADQASAPAEFAALQEAIDLTAQDVPAWLRNLSLBARLQSGS KLSTHADLTHDQAIQRYQTFVAEVLPAQERASQALTEKLAIVPGVPGPFPALNR RFQQAALFRFANVDLGVTHNQSHSVITGNQVGLQCELTLPQAKQGGSPDRA QREAAWALSADNAIAPELDVMDLIRTRQLAHNADENFRWKELEDRVDYTP EQCLDFVSRVDEVPVLLSGVLGDIARQLGLESRPWDYNNRNLDPGEGESLRFPT GADLEALQESLDLPDLAARFQMRSEGLLDLESRPGKWTAYCNYPTTNEPFL MNVGTAEDEVVLVEHGAHPGFGYSGDRQPLVNNRWSPIEFVEIPSMAEFTLDEL SHVSDDLARVRSKQLEGVITAFIPMAAQDAFQHWLYAEPEDVTIADLAKWLELD RTFHFFDQKSLGDSVRKAGWYVHIQVQPYEYEVAMCYLAARVAIVRAARQDPAAL GRYKDALGNVTGVPELYRAAGAEFFREYINGLMFLKEQLG"  
/complement (2005..3645)  
/gene="DRA0207"  
/complement (2005..3645)  
/gene="DRA0207"  
/feature="similar to PID:1653945 percent identity: 49.28; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI2422.1"  
/db\_xref="GI:6460717"  
/translation="MNFEEVLSFRLSSVAHIALNRASLKEEMSKRSLSALLGALL SACSPENGPQAPQLPFPTEAPGTPSGKRGVLWDVAFGPGIKTPEEVSTLUDTAR AMVNTLVQVGRGDCYCNNAAPRTDDPEVPAGDFADVIKAKAHADIEVNAII TADVNSVTTPKSAHEAFNTHGPNAGRDNLVTKADGTAKAGADWLDPGHDAAY TRNMYVSAKAYVDGIMEDRVRYPDYNEUGVQWQCYNETALRYAEATGATGDDP ADPQASNRQGVNLR

RAIRQAVEGAEDDEAAGRDVQAVNIGPDDVLIAVAASGTTTPVILGAARSRAAGALT  
VGLANPGLAIAVDCPVLDTGPEIISGSTRKAGTAQKIALNTLSALMVLGKL  
YGNLVVDVTRATNAKLEDRRLVQVHATGADADAQAALSECAGSVKLTALVMLKGLGA  
QGAORLEGAGGHARQVLGEAGLGTSA"  
complement(10796..11398)  
/gene="DRA0214"  
complement(10796..11398)  
/note="similar to GB:M99166 SP:P30849 PID:148264 GB:U00096  
PID:1651499 percent identity: 54.36; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="trp repressor binding protein wrbA, putative"  
/protein\_id="AA012417.1"  
/db\_xref="GI:6460712"  
/translation="MTAPVKLAIFVYSTGTGYAVQAQAAEAGRAAGAEVRLKVRRT  
APQVIDGDQANKANIEAKDVEATPADLEWAEIAIVFSSPTRFGATSQMRAFDITL  
GLWSSGKLANTFSAITSAONVNGOETLLOTLYMTAMHGAULTPPGYTDEVIFKS  
GNGPYGASVTANGQPLENDRASIRHVRQVELTAKLLG"  
11664..12059  
/gene="DRA0215"  
11664..12059  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA012499.1"  
/db\_xref="GI:6460794"  
/translation="MGGLHPRQLGHLRSLDELPMFGFWLNPREVAFLKAGLNAAG  
LTGMGWTGPHALGPAPFAALLSNEQARLDGLVRLVAHHLTCLTLGAPLLDPND  
TERVQRIEATRPPLFVDEVMAGREDEGA"  
12068..13069  
/gene="DRA0216"  
12068..13069  
/note="DRA0216"  
/gene="DRA0216"

Alignment Scores:  
Pred. No.: 4e+04 Length: 198677  
Score: 88.50 Matches: 37  
Percent Similarity: 49.59% Conservative: 24  
Best Local Similarity: 30.08% Mismatches: 36  
Query Match: 12.50% Indels: 26  
DB: 1 Gaps: 8

US-10-087-573-2 (1-141) x AE001863 (1-198677)

QY 25 MetArgGluAlaLeuLeu-----ArgValLysSerSerGluArgLeuAla 39  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 147384 CTGCGTAGCGGTGCTGAGCTGCATGCGCGCGCATCGAAACCGGCTCAGAGTCT 147443

QY 40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGly----- 57  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 147444 -----CAGGCCCTG-----CGCGGATTCTCGCGGAGAGGACGGTG 147482

QY 58 -----AlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLys 75  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 147483 ACGTCAAGCGCGCGCGCGCTGCGCGGTCTCACGCCCAATCTCAGCGGTGCGT 147542

QY 76 Pro-----ProArgProGlnSerThrLysSerProGluLeu-----ArgGluLeuSer 91  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 147543 CCAAGTCCAGCGCGCTACTCCGATCGCATCGCAACCGCAACCGAGCGGTGGAA 147602

QY 92 ArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArg 111  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 147603 AAACGGCTCCGGAGCTGGAAAGAAGTCCGCAAGGAGCGCGGAGCGATCCCGCGGC 147662

QY 112 LeuPro-GluGlyHisProLeu-----LeuGluLysArgAlaGluTyrPheAr 127  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 147663 ---CCTGCACCGGCACCGCTCCATTCCTCATCCCTGCGCAGCGGCCCGCGCTCGG 147719

QY 127 ghisLeu 129

Db 147720 CCCCCTC 147726

RESULT 34  
LOCUS BD180073/c 846 bp DNA linear PAT 15-MAY-2003  
DEFINITION Highly thermophilic bacterium-derived protein and gene encoding it.  
ACCESSION BD180073  
VERSION BD180073.1 GI:30790991  
KEYWORDS JP 2002325574-A/564.  
SOURCE Thermus thermophilus  
ORGANISM Thermus thermophilus  
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
Thermus.

REFERENCE  
1 (Bases 1 to 846)  
Kuramitsu,N. and Yokoyama,S.  
AUTHORS  
Highly thermophilic bacterium-derived protein and gene encoding it  
TITLE  
Patent: JP 2002325574-A 564 12-NOV-2002;  
JOURNAL THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH  
COMMENT OS Thermus thermophilus  
PN JP 2002325574-A/564  
PD 12-NOV-2002  
PF 23-FEB-2001 JP 2001116171  
PI NARUKI KURAMITSU,SHIGEYUKI YOKOYAMA  
PC C12N15/09,C12N15/09,C07K14/195,C12N1/15,C12N1/19,C12N1/21,PC  
C12N5/10,  
PC C12N9/88,C12P21/02/(C12N9/88,C12R1:01),(C12N15/09,C12R1:01),  
PC (C12P21/02,C12R1:01),(C12N15/00,C12N15/00,C12N5/00,C12N15/00,  
PC C12R1:01)  
CC Highly thermophilic bacterium-derived protein and gene CC  
encoding it

FEATURES  
source  
1..846 Location/Qualifiers  
CDS 89 a 324 c 283 g 150 t  
BASE COUNT 89 a 324 c 283 g 150 t  
ORIGIN

Alignment Scores:  
Pred. No.: 71.4 Length: 846  
Score: 88.00 Matches: 34  
Percent Similarity: 49.18% Conservative: 26  
Best Local Similarity: 27.87% Mismatches: 48  
Query Match: 12.43% Indels: 14  
DB: 6 Gaps: 6

US-10-087-573-2 (1-141) x BD180073 (1-846)

QY 25 MetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeu 44  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 398 CTTTCCAGGGCGGCTCCCGCACCTCTCCGCGAGAGGTCTTCGGTCTCCGCTCAGC 339

QY 45 AlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer-----AlaIleAla 62  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 338 -----ARGCGGAAAGCGGCTTCTCCCGCTCAGAGGCTCTCCAGGCGCTTCTCC 285

QY 63 AlaThrValThrProLysGlyAlaSerMetLysLeuLys-----ProArgPro 79  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 284 ACCCAAAGACCTCCAAAGAGGAGCATCTCAGGCTGAAGAGGCTTCAGGAGGAGGCC 225

QY 80 GlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLys 99  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 224 AGGACCAACC---CACCCCAAGCGGCTCAGGAGGAGTCCAGGAGAAAGCGCGGC 168

QY 100 ThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisPro----- 117  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 167 AGGGGAAGAGGAGGCGCCCAAGAGCCAGTAGAGCCCGCGAGGAAACCCCGGACC 108

QY 118 LeuLeuGlnLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyVal 137  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```
Db 107 CTCTCCAGGCCCGCAGGATAC-----CGCCCCATCCAGACCACGGGGTGG 60
QY 138 AsnArg 139
Db 59 AGCGG 54

RESULT 35
AP005426 148797 bp DNA linear HTG 13-JUN-2002
LOCUS Oryza sativa (japonica cultivar-group) chromosome 9 clone P0668D04,
DEFINITION *** SEQUENCING IN PROGRESS ***
ACCESSION AP005426
VERSION AP005426.1 GI:21396536
KEYWORDS HTG; HTGS PHASE2.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 Sasaki,T., Matsumoto,T. and Katayose,Y.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 9, PAC
Clone:P0668D04
Published Only in Database (2002)
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 148797)
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasak@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..148797
location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="9"
/clone="P0668D04"
BASE COUNT 42512 a 32195 c 31612 g 42428 t 50 others
ORIGIN

Alignment Scores:
Pred. No.: 3.14e+04 Length: 148797
Score: 88.00 Matches: 39
Percent Similarity: 39.16% Conservative: 26
Best Local Similarity: 23.49% Mismatches: 63
Query Match: 12.43% Indels: 38
DB: 2 Gaps: 5

US-10-087-573-2 (1-141) x AP005426 (1-148797)
QY 2 GluSerThrSerThrThrThrAsn-----PheValalaGluAsnArgProThrPheGly 19
Db 145582 GAGAAACACGACGACGCGGAAGTACACTGACAGTATAATACACAGTAGACT----- 145632
QY 20 GluThrPheAspValMetArgGluAlaLeuLeuArgValIysSerSerGluArgLeuAla 39
Db 145633 -----TGTCAGCGAGGAGGCATGCATGGAAGCTCTAGCACAGGTAGTATCGA 145680
QY 40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59
```



REFERENCE	Streptomycineae; Streptomycetaceae; Streptomyces.
AUTHORS	1 Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Krieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wierzbicki, A., Woodward, J., Barrall, B.G., Parkhill, J. and Hopwood, D.A.
TITLE	Complete genome sequence of the model actinomycete Streptomycetes coelicolor A3(2)
JOURNAL	Nature 417 (6885), 141-147 (2002)
MEDLINE	21996410
PUBMED	12000953
REFERENCE	2 (bases 1 to 283100)
AUTHORS	Bentley, S.D.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-2002) Submitted on behalf of the Streptomycetes sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk
COMMENT	On or before Oct 26, 2002 this sequence version replaced gi:20520819, gi:20520879, gi:20520880, gi:20520919, gi:20520821, gi:20520887, gi:20520898, gi:20520899, gi:20520884, gi:20520704, gi:20520915, gi:20520918.
FEATURES	Location/Qualifiers
source	1. 283100 /organism="Streptomycetes coelicolor A3(2)" /mol_type="genomic DNA" /strain="A3(2)" /db_xref="taxon:100226" 59..520 /gene="SCO1698" /note="synonym: SCI30A.19" 59..520 /gene="SCO1699" /note="SCI30A.19, unknown, len: 153 aa; similar to many of undefined function e.g. TR:P98807 (EMBL:292770) Hypothetical protein from Mycobacterium tuberculosis (151 aa) fasta scores; opt: 639, z-score: 785.9, E(): 0, (60.8% identity in 148 aa overlap) and SW:NODN_RHIV nodulation protein from Rhizobium leguminosarum (161 aa) fasta scores; opt: 388, z-score: 482.2, E(): 1.6e-19, (44.8% identity in 145 aa overlap)" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="CAB46796.1" /db_xref="GI:5441772"
gene	/translation="MAGPRIIPASADVEKAAVGEQLGYTDMLEVDQKRIDLFAEATGDH OWIHVDPRRAAGPFGTIAHGTYLTSLPLFGPOLIRVGVKMGVNTKVPFSP VPVSGRLRATVITGVEDVKGIIQVSVAFTREREKGGKPKVCAESVRYL" complement(511..1158) /gene="SCO1699" /note="synonym: SCI30A.20c" complement(511..1159) /gene="SCO1699"
CDS	/note="SCI30A.20c, possible transcriptional regulator, len: 215 aa; similar to many e.g. TR:CAB39717 (EMBL:AL049485) putative regulatory protein from Streptomycetes coelicolor (223 aa) fasta scores; opt: 330, z-score: 413.5, E(): 1.1e-15, (33.8% identity in 216 aa overlap) and TR:ADJ356 (EMBL:AF080235) Lank regulator of landomycin biosynthesis in Streptomycetes cyanogenus S136 (192 aa) fasta scores; opt: 170, z-score: 218.5, E(): 7.6e-05, (31.8% identity in 154 aa overlap). Contains Pfam match to entry PF00440 tetr, Bacterial regulatory proteins, tetr family. Contains possible helix-turn-helix motif (-4.32 SD) 43-64aa." /codon_start=1 /transl_table=11 /product="putative transcriptional regulator"
gene	/misc_feature complement(1636..2748) /gene="SCO1701" /note="Pfam match to entry PF00441 Acyl-CoA dh, Acyl-CoA dehydrogenase, score 565.20, E-value 4.3e-166." 2850..3491 /gene="SCO1702" /note="synonym: SCI30A.23" 2850..3491 /gene="SCO1702" /note="SCI30A.23, possible transcriptional regulator, len: 213 aa; similar to many e.g. TR:Q53901 (EMBL:M64683) regulatory protein from Streptomyces coelicolor (259 aa) fasta scores; opt: 227, z-score: 273.2, E(): 6.9e-08, (32.3% identity in 217 aa overlap). Contains Pfam match to entry PF00440 tetr, Bacteria regulatory proteins, tetr family and Prosite match to F501081 Bacterial regulatory proteins, tetr family signature. Contains possible helix-turn-helix motif (+5.17 SD) 31-52aa." /codon_start=1 /transl_table=11
CDS	/misc_feature complement(1509..2760) /gene="SCO1701" /note="synonym: SCI30A.22c" complement(1509..2760) /gene="SCO1701" /note="SCI30A.22c, probable acyl-CoA dehydrogenase, len: 383 aa; similar to many both prokaryote and eukaryote eg. SW:ACDB_BACSU acyl-CoA dehydrogenase from Bacillus subtilis (379 aa) fasta scores; opt: 109.1 z-score: 1197.8, E(): 0, (43.8% identity in 377 aa overlap) and SW:ACDS_RAT acyl-CoA dehydrogenase from Rattus norvegicus (Rat) (412 aa) fasta scores; opt: 1079, z-score: 1184.2, E(): 0, (44.6% identity in 372 aa overlap). Contains Pfam match to entry PF00441 Acyl-CoA dh, Acyl-CoA dehydrogenase." /codon_start=1 /transl_table=11 /product="putative acyl-CoA dehydrogenase" /protein_id="CAB46799.1" /db_xref="GI:5441775"
gene	/misc_feature complement(1636..2748) /gene="SCO1701" /note="Pfam match to entry PF00441 Acyl-CoA dh, Acyl-CoA dehydrogenase, score 565.20, E-value 4.3e-166." 2850..3491 /gene="SCO1702" /note="synonym: SCI30A.23" 2850..3491 /gene="SCO1702" /note="SCI30A.23, possible transcriptional regulator, len: 213 aa; similar to many e.g. TR:Q53901 (EMBL:M64683) regulatory protein from Streptomyces coelicolor (259 aa) fasta scores; opt: 227, z-score: 273.2, E(): 6.9e-08, (32.3% identity in 217 aa overlap). Contains Pfam match to entry PF00440 tetr, Bacteria regulatory proteins, tetr family and Prosite match to F501081 Bacterial regulatory proteins, tetr family signature. Contains possible helix-turn-helix motif (+5.17 SD) 31-52aa." /codon_start=1 /transl_table=11
CDS	/misc_feature complement(1509..2760) /gene="SCO1701" /note="synonym: SCI30A.22c" complement(1509..2760) /gene="SCO1701" /note="SCI30A.22c, probable acyl-CoA dehydrogenase, len: 383 aa; similar to many both prokaryote and eukaryote eg. SW:ACDB_BACSU acyl-CoA dehydrogenase from Bacillus subtilis (379 aa) fasta scores; opt: 109.1 z-score: 1197.8, E(): 0, (43.8% identity in 377 aa overlap) and SW:ACDS_RAT acyl-CoA dehydrogenase from Rattus norvegicus (Rat) (412 aa) fasta scores; opt: 1079, z-score: 1184.2, E(): 0, (44.6% identity in 372 aa overlap). Contains Pfam match to entry PF00441 Acyl-CoA dh, Acyl-CoA dehydrogenase." /codon_start=1 /transl_table=11 /product="putative acyl-CoA dehydrogenase" /protein_id="CAB46799.1" /db_xref="GI:5441775"
gene	/misc_feature complement(1636..2748) /gene="SCO1701" /note="Pfam match to entry PF00441 Acyl-CoA dh, Acyl-CoA dehydrogenase, score 565.20, E-value 4.3e-166." 2850..3491 /gene="SCO1702" /note="synonym: SCI30A.23" 2850..3491 /gene="SCO1702" /note="SCI30A.23, possible transcriptional regulator, len: 213 aa; similar to many e.g. TR:Q53901 (EMBL:M64683) regulatory protein from Streptomyces coelicolor (259 aa) fasta scores; opt: 227, z-score: 273.2, E(): 6.9e-08, (32.3% identity in 217 aa overlap). Contains Pfam match to entry PF00440 tetr, Bacteria regulatory proteins, tetr family and Prosite match to F501081 Bacterial regulatory proteins, tetr family signature. Contains possible helix-turn-helix motif (+5.17 SD) 31-52aa." /codon_start=1 /transl_table=11
CDS	/misc_feature complement(1509..2760) /gene="SCO1701" /note="synonym: SCI30A.22c" complement(1509..2760) /gene="SCO1701" /note="SCI30A.22c, probable acyl-CoA dehydrogenase, len: 383 aa; similar to many both prokaryote and eukaryote eg. SW:ACDB_BACSU acyl-CoA dehydrogenase from Bacillus subtilis (379 aa) fasta scores; opt: 109.1 z-score: 1197.8, E(): 0, (43.8% identity in 377 aa overlap) and SW:ACDS_RAT acyl-CoA dehydrogenase from Rattus norvegicus (Rat) (412 aa) fasta scores; opt: 1079, z-score: 1184.2, E(): 0, (44.6% identity in 372 aa overlap). Contains Pfam match to entry PF00441 Acyl-CoA dh, Acyl-CoA dehydrogenase." /codon_start=1 /transl_table=11 /product="putative acyl-CoA dehydrogenase" /protein_id="CAB46799.1" /db_xref="GI:5441775"
gene	/misc_feature complement(1636..2748) /gene="SCO1701" /note="Pfam match to entry PF00441 Acyl-CoA dh, Acyl-CoA dehydrogenase, score 565.20, E-value 4.3e-166." 2850..3491 /gene="SCO1702" /note="synonym: SCI30A.23" 2850..3491 /gene="SCO1702" /note="SCI30A.23, possible transcriptional regulator, len: 213 aa; similar to many e.g. TR:Q53901 (EMBL:M64683) regulatory protein from Streptomyces coelicolor (259 aa) fasta scores; opt: 227, z-score: 273.2, E(): 6.9e-08, (32.3% identity in 217 aa overlap). Contains Pfam match to entry PF00440 tetr, Bacteria regulatory proteins, tetr family and Prosite match to F501081 Bacterial regulatory proteins, tetr family signature. Contains possible helix-turn-helix motif (+5.17 SD) 31-52aa." /codon_start=1 /transl_table=11
CDS	/misc_feature complement(1509..2760) /gene="

REFERENCE	Streptomycineae; Streptomycetaceae; Streptomyces.
AUTHORS	Bentley, S.D., Chater, K.F., Cerdano-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Krieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wierzbicki, A., Woodward, J., Barrall, B.G., Parkhill, J. and Hopwood, D.A.
TITLE	Complete genome sequence of the model actinomycete Streptomycetes coelicolor A3(2)
JOURNAL	Nature 417 (6885), 141-147 (2002)
MEDLINE	21996410
PUBMED	12000953
REFERENCE	2 (bases 1 to 283100)
AUTHORS	Bentley, S.D.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-2002) Submitted on behalf of the Streptomycetes sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk
COMMENT	On or before Oct 26, 2002 this sequence version replaced gi:20520819, gi:20520879, gi:20520880, gi:20520919, gi:20520821, gi:20520887, gi:20520898, gi:20520899, gi:20520884, gi:20520704, gi:20520915, gi:20520918.
FEATURES	Location/Qualifiers
source	1. 283100
	/organism="Streptomycetes coelicolor A3(2)"
	/mol_type="genomic DNA"
	/strain="A3(2)"
	/db_xref="taxon:100226"
gene	59..520
	/gene="SCO1698"
	/note="synonym: SCI30A.19"
CDS	59..520
	/gene="SCO1699"
	/note="SCI30A.19, unknown, len: 153 aa; similar to many of undefined function e.g. TR:P98807 (EMBL:292770) Hypothetical protein from Mycobacterium tuberculosis (151 aa) fasta scores; opt: 639, z-score: 785.9, E(): 0, (60.8% identity in 148 aa overlap) and SW:NODN_RHIV nodulation protein from Rhizobium leguminosarum (161 aa) fasta scores; opt: 388, z-score: 482.2, E(): 1.6e-19, (44.8% identity in 145 aa overlap)"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="CAB46796.1"
	/db_xref="GI:5441772"
	/translation="MAGPRIPASADVEKAAVGEQLGYTDMLEVDQKRIDLFAEATGDH OWIHVDPRRAAGPFGTIAHGUYLTLLPLFGPOLIRVGVKMGVNTKVPFSP VPVSGRLRATVITGVEDVKGIIQVSVAFTREREKGGKPKVCAESVRYLL"
gene	complement(511..1158)
	/gene="SCO1699"
	/note="synonym: SCI30A.20c"
CDS	complement(511..1158)
	/gene="SCO1699"
	/note="SCI30A.20c, possible transcriptional regulator, len: 215 aa; similar to many e.g. TR:CAB39717 (EMBL:AL049485) putative regulatory protein from Streptomycetes coelicolor (223 aa) fasta scores; opt: 330, z-score: 413.5, E(): 1.1e-15, (33.8% identity in 216 aa overlap) and TR:ADJ356 (EMBL:AF080235) Lank regulator of landomycin biosynthesis in Streptomycetes cyanogenus S136 (192 aa) fasta scores; opt: 170, z-score: 218.5, E(): 7.6e-05, (31.8% identity in 154 aa overlap). Contains Pfam match to entry PF00440 tetr, Bacterial regulatory proteins, tetr family. Contains possible helix-turn-helix motif (-4.32 SD) 43-64aa."
	/codon_start=1
	/transl_table=11
	/product="putative transcriptional regulator"
	/protein_id="CAB46797.1"
	/db_xref="GI:5441773"
	/translation="MSTAAETGGDIEPHEVTPDARLLVAEFAERGVHATTI RTADRAGSPAAIYHYTKELHRIIRIGHTRAVAILRQAAQEGEAAERLADAY SSFVRHAGRTARVWVYELDALGEAREEILTLRQCDAAVRGIIIDGVAAGEDFY PDVKGTTAVLSLCIDVARFWNPGRTPDEVGALYADLVLRMVGAEPSGATAQR"
misc_feature	complement(943..1083)
	/gene="SCO1699"
	/note="Pfam match to entry PF00440 tetr, Bacterial regulatory proteins, tetr family, score 60.70, E-value 4e-15."
gene	complement(1219..1533)
	/gene="SCO1700"
	/note="synonym: SCI30A.21c"
CDS	complement(1219..1533)
	/gene="SCO1700"
	/note="SCI30A.21c, putative membrane protein, len: 104 aa. Contains possible hydrophobic membrane spanning regions"
	/codon_start=1
	/transl_table=11
	/product="putative membrane protein"
	/protein_id="CAB46798.1"
	/db_xref="GI:5441774"
	/translation="MSDTPVKQNTAAFYGOAVASFVAMAATAIGIYQLOADAWVRA FLATAVLYVTSFTLAKIIRDRQEVQIVSRVQDARLRLVVEHDPFKPGTTPAG RP"
gene	complement(1509..2760)
	/gene="SCO1701"
	/note="synonym: SCI30A.22c"
CDS	complement(1509..2760)
	/gene="SCO1701"
	/note="SCI30A.22c, probable acyl-CoA dehydrogenase, len: 383 aa; similar to many both prokaryote and eukaryote eg. SW:ACDB_BACSU acyl-CoA dehydrogenase from Bacillus subtilis (379 aa) fasta scores; opt: 109.1, z-score: 1197.8, E(): 0, (43.8% identity in 377 aa overlap) and SW:ACDS_RAT acyl-CoA dehydrogenase from Rattus norvegicus (Rat) (412 aa) fasta scores; opt: 1079, z-score: 1184.2, E(): 0, (44.6% identity in 372 aa overlap). Contains Pfam match to entry PF00441 Acyl-CoA_dh, Acyl-CoA dehydrogenase."
	/codon_start=1
	/transl_table=11
	/product="putative acyl-CoA dehydrogenase"
	/protein_id="CAB46799.1"
	/db_xref="GI:5441775"
	/translation="MNLSEEQTAVRQLARDFEREIAPHVVMEDRAEEVDRSLVKY LGEVFLGLTIDQYGGSGDHLAYCLVTELRGDSVVRGIVSGIVAKTIAAAG DEOKRWLPSTGDEYGVGFCFTEPTGSDAGNLITRAVRDGGVAVNGTKMFINFG TWADVLLPRLTSGVHQGVSAFLVPTDTPGTITRTHKGLGRQATAEILGVLEVR VPASMLAPRGKGFVSAMSLAKGRMSVAGCVGIAQALDAVRVAGREGEFGKTLA HQLVQELISDIALDVAARLLTWVADLIDRQPPAVFSSKAKLPASAAVRAANNA HQVFGAYGIDIEPAGKLLARDAKRWMLTYEGTSQIQKLVIGRALTGVSAF"
misc_feature	complement(1636..2748)
	/gene="SCO1701"
	/note="Pfam match to entry PF00441 Acyl-CoA_dh, Acyl-CoA dehydrogenase, score 565.20, E-value 4.3e-166."
gene	2850..3491
	/gene="SCO1702"
	/note="synonym: SCI30A.23"
CDS	2850..3491
	/gene="SCO1702"
	/note="SCI30A.23, possible transcriptional regulator, len: 213 aa; similar to many e.g. TR:Q53901 (EMBL:M64683) regulatory protein from Streptomyces coelicolor (259 aa) fasta scores; opt: 227, z-score: 273.2, E(): 6.9e-08, (32.3% identity in 217 aa overlap). Contains Pfam match to entry PF00440 tetr, Bacteria regulatory proteins, tetr family and Prosite match to F501081 Bacterial regulatory proteins, tetr family signature. Contains possible helix-turn-helix motif (+5.17 SD) 31-52aa."
	/codon_start=1
	/transl_table=11



```

/product="putative transcriptional regulator"
/db_xref="GI:5411776"
/translation="WARPRLKLLSTDRIVETARALVDAEGLAAVSTRLAELGVSGP
SLYNHFRKDETLAVDSQVLDLMEFDGRTALHDAVSYPRALRDHNIVP
VLHAGFGRPALHADIYGMVPCAPQAQASIGALMRYFYVMSALGSPAGFVD
DGAADPADIPHLQAHLLAEQCEKIDERAFETGLTALDGLAQQAQVAVQGV"
2889..3029
/gene="SC01702"
/notes="pfam match to entry PF00440 tetR, Bacterial
regulatory proteins, tetR family, score 45.00, E-value
9.2e-11."
2925..3017
/gene="SC01702"
/notes="PS01081 Bacterial regulatory proteins, tetR family
signature."
3555..4268
/gene="SC01703"
/notes="synonym: SC130A.24"
3555..4268
/gene="SC01703"
/notes="SC130A.24, possible transcriptional regulator, len:
237aa; similar to many eg. SW:MERR_STRLI probable mercury
resistance operon repressor from Streptomyces lividans
(125 aa) fasta scores; opt: 144, z-score: 183.3, E():
0.007, (38.2% identity in 89 aa overlap). Contains Pfam
match to entry PF01022 HTH_5, Bacterial regulatory
protein, arsr family. Contains possible helix-turn-helix
motif (43.92 SD) 37-58aa."
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB46801.1"
/db_xref="GI:5411777"
/translation="WTEKDAGPGLARLGLADDETRACALLLDGRWTAGSLARH
AGVASTUSEHLKVLGGLLAERQGRHVRMADRLVHLVEDLAQVAPAAARR
PRLTASAGSAMAGRTICYDHLAAGRLGIADVTAGLLRQDTGFTADGLGWF
SAGILPRTRRPLAALCTTERRPHLAGVAGRALCRHALDTGNCVRIGSERAVRT
PAGEDALSOLDIDPAALR"
3597..3833
/gene="SC01703"
/notes="pfam match to entry PF01022 HTH_5, Bacterial
regulatory protein, arsr family, score 55.80, E-value
9.1e-13."
4342..5316
misc_feature
gene

Alignment Scores:
Pred. NO.: 6.67e-04 Length: 283100
Score: 88.00 Matches: 44
Percent Similarity: 43.56% Conservative: 27
Best Local Similarity: 26.93% Mismatches: 58
Query Match: 12.43% Indels: 35
DB: 1 Gaps: 10

US-10-087-573-2 (1-141) x SC0939110 (1-283100)

QY 1 MetGluSerThrSerThrThrThrPheValAlaGluAsnArgProThrPheGlyGlu 20
Db 261762 ATCTCCACGCTCTCTGCACCAACACGCGGCGCGATGGCCAGAGTCTCGACGAG 261703
QY 21 ThrPheAspValMetArgGluAlaLeuLeuArgVal-----LysSerSerGluArg 37
Db 261702 AACTTCGGCATCGTCAAGGCGCTGATGACGAGCGTCCACGCGGTACACGACGACGCGC 261643
QY 38 LeuAla-----MetLeuArgAlaLeuAlaGlyMetCysGlyHisArg 51
Db 261642 ATCTGGACTTCCCGACCAAGACCTGCGCGCGCGCGTGC-----GCCGCGGAGAAC 261589
QY 52 ValLeuPro---GlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLys----- 68
Db 261588 ATCATCCGACACACACGCGCGCGCCCAAGGCCCGCTGGTCTCTCCGACGAGTCAAG 261529
QY 69 -----GlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSer 84

```

```

Db 261528 GGCAAGCTCGACGCGATCGCGATCGCGGTC-----CCGGTCCCGACGGTCCGCCACC 261475
QY 85 ProGluLeuArgGluLeuSerArgLysIleArg-----GluMetAsnLysThrIleSer 102
Db 261474 GACCTGCTGTGGAGCTCCAGCGGAGGTCCACGACGACGAGGTCAACGC-CGCGTT--- 261419
QY 103 GlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLys--- 121
Db 261418 CAGAAGAGCGCCCGA-----CGACGCGGACCTCAAGGGCATCTCTGTCTACACCGAGGA 261365
QY 122 -----ArgAlaGluTyrrPheargHisLeuArgSerLeuLys 133
Db 261364 CGCGATCGTCTCTCGACATCACCGCGGCGCTCTGCGACCTCGATCTCCTCCT 261305
QY 134 SerGlnGly 136
Db 261304 GACCATGGT 261296

RESULT 39
LOCUS AC125900
DEFINITION Rattus norvegicus clone CH230-319G17, *** SEQUENCING IN PROGRESS
***, 12 unordered pieces.
ACCESSION AC125900.3 GI:25008624
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 300813)
Muzny, D., Maric, M., Metzger, M., Lee, S., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderson, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devita, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guetara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulvy, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kuwis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshehwa, L., Loulseg, H., Lozano, R., Lu, X., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, P., Polindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzos, M., Quirz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Rilly, B., Reilly, M., Ren, I., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,

```





Db 160195 CAAGTGCTCCGTTACAGTCCCAATCCCTAGCCTTGATCTGCTGCTTTTAGATTA 160254  
 Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 Db 160255 GAACCAAGGGAGTGAAGCGCTG-----CCAGTCCTTGGCACCTTG----- 160296  
 Qy 95 ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGlu 114  
 Db 160297 -----CGGTGTCATCTCCACCGTCAAGTATCTCACTATGGCCGAAG 160341  
 Qy 115 GlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeu 132  
 Db 160342 GGTATCTCAAAGCAGCAAGCAAGAACCACTGTGGCTGCTATCTCAGGCCACTA 160395  
 RESULT 40  
 AC118450  
 LOCUS Rattus norvegicus clone CH230-144P6, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION  
 \*\*\* 35 unordered pieces.  
 AC118450  
 AC118450.3 GI:23908096  
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 308821)  
 Muzny,D.,Maric,E., Metzker,M.,Lee., Abramzon,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
 Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biewalo,K., Blair,J., Blankenbush,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Ceasat,H., Center,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Gante,R., Garcia,A., Garner,T., Garza,M.,  
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,B., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensuhewa,L., Louissegh,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankerville,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwakeleneh,O., Okwuonu,G., Olarnpusagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
 Piazon,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Riley,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,I., Rojars,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soza,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera V. Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,

Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 308821)  
 Worley,K.C.  
 Direct Submission  
 Submitted (17-APR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 308821)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (12-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Oct 12, 2002 this sequence version replaced gi:21903188.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GWHK  
 Center clone name: CH230-144P6  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 0 bases at least Q40  
 Consensus quality: 0 bases at least Q30  
 Consensus quality: 0 bases at least Q20  
 Estimated insert size: 268894; sum-of-contigs estimation  
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
 \* NOTE: This sequence may represent more than one clone.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 35 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1022: contig of 1022 bp in length  
 \* 1023 1122: gap of unknown length  
 \* 1123 2637: contig of 1515 bp in length  
 \* 2638 2737: gap of unknown length  
 \* 2738 4675: contig of 1937 bp in length  
 \* 4675 4774: gap of unknown length  
 \* 4775 6221: contig of 1447 bp in length  
 \* 6222 6321: gap of unknown length  
 \* 6322 7397: contig of 1076 bp in length  
 \* 7398 7497: gap of unknown length  
 \* 7498 8902: contig of 1305 bp in length  
 \* 8903 13598: contig of 4696 bp in length  
 \* 13599 13698: gap of unknown length  
 \* 13699 15957: contig of 2259 bp in length

\* 15958 16057: gap of unknown length  
\* 16058 21808: contig of 5751 bp in length  
\* 21809 21808: gap of unknown length  
\* 21909 21808: contig of 1322 bp in length  
\* 22301 21808: gap of unknown length  
\* 22331 21808: contig of 1352 bp in length  
\* 24682 21808: gap of unknown length  
\* 24782 21808: contig of 3720 bp in length  
\* 24783 21808: gap of unknown length  
\* 28503 21808: contig of 74248 bp in length  
\* 28603 102850: contig of 93331 bp in length  
\* 102851 196281: contig of 93331 bp in length  
\* 196282 196381: gap of unknown length  
\* 196382 197669: contig of 1288 bp in length  
\* 197670 197769: gap of unknown length  
\* 197770 199047: contig of 1278 bp in length  
\* 199048 200429: contig of 1282 bp in length  
\* 200430 200529: gap of unknown length  
\* 200530 234122: contig of 33593 bp in length  
\* 234123 234222: gap of unknown length  
\* 234223 234931: contig of 1269 bp in length  
\* 234932 235591: gap of unknown length  
\* 235592 238938: contig of 3347 bp in length  
\* 238939 239038: gap of unknown length  
\* 239039 240391: contig of 1353 bp in length  
\* 240392 240491: gap of unknown length  
\* 240492 243023: contig of 2532 bp in length  
\* 243024 243123: gap of unknown length  
\* 243124 244444: contig of 1321 bp in length  
\* 244445 248267: contig of 3723 bp in length  
\* 248268 248367: gap of unknown length  
\* 248368 249563: contig of 1196 bp in length  
\* 249564 251245: contig of 1196 bp in length  
\* 251246 251345: gap of unknown length  
\* 251346 253225: contig of 1880 bp in length  
\* 253226 253325: gap of unknown length  
\* 253326 257369: contig of 4044 bp in length  
\* 257370 257469: gap of unknown length  
\* 257470 259280: contig of 1811 bp in length  
\* 259281 259380: gap of unknown length  
\* 259381 261051: contig of 1671 bp in length  
\* 261052 263987: contig of 2836 bp in length  
\* 263988 264087: gap of unknown length  
\* 264088 267533: contig of 3446 bp in length  
\* 267534 267633: gap of unknown length  
\* 267634 279620: contig of 11987 bp in length  
\* 279621 279720: gap of unknown length  
\* 279721 298462: contig of 18742 bp in length  
\* 298463 298562: gap of unknown length  
\* 298563 308821: contig of 10259 bp in length.

FEATURES  
    source  
        1..308821  
            /organism="Rattus norvegicus"  
            /mol\_type="genomic DNA"  
            /db\_xref="taxon:10116"  
            /clone="CH230-144P6"  
            1..1022  
                /note="wgs\_end\_extension  
                clone\_end:T7"  
                2080..2962  
                /note="clone\_boundary  
                clone\_end:T7  
                site:ECORI  
                end\_sequence:RWAXX87JB"  
                2738..4674  
                /note="wgs\_contig"

misc\_feature  
misc\_feature  
misc\_feature

Alignment Scores:  
Pred. No.: 7.38e+04 Length: 308821

Score: 88.00 Matches: 33  
Percent Similarity: 42.37% Conservative: 17  
Best Local Similarity: 27.97% Mismatches: 44  
Query Match: 12.43% Indels: 24  
DB: 2 Gaps: 4

US-10-087-573-2 (1-141) x AC118450 (1-308821)  
QY 24 ValMetArgGluAlaLeuLeuArgValLysSerGluArgLeuAlaMetLeuArgAla 43  
Db 39704 GTAATAAATACTCCCTAATAAAGAAATATATAGAGAAATGCAACTTATTGCTTC 39763  
QY 44 LeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaLeuAla 63  
Db 39764 CTACAA-----CAGCAGGGCTCTCCCGGGTATGGCTGCACAGCCAGGCACAG 39811  
QY 64 ThrValThrPro-----LysGlyAlaSerMetLysLeu 74  
Db 39812 CAAGTGCCTCCCTTACAGTCCCATTCCTAGCCTTGATCTGCTCTCTTTAGATTA 39871  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 39872 GAACCAAGGGGAGTGAAGCCTG-----CCAGTCTTGGCACCCTTG----- 39913  
QY 95 ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGlu 114  
Db 39914 -----CGGGTGTCATCTCCACGGTCAAGTATCTCCTATGCGCCAGAAAG 39958  
QY 115 GlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeu 132  
Db 39959 GGTCAATCCAAAGCAGGCAAGAACCACTGTGGCTGGTCACTCAGGCCACTA 40012

Search completed: November 17, 2003, 13:45:56  
Job time : 2504 secs



FT		/*tag= a	
FT		/product= "15kd protein"	
XX			
PN	EP1238983-A1.		
PD	11-SEP-2002.		
PP	04-MAR-2002; 2002EP-0075930.		
PP	06-MAR-2001; 2001EP-0200816.		
XX	(ALKU ) AKZO NOBEL NV.		
XX	Schettlers TPW, Carcy BPD, Drakulovski PR, Gorenflot AF;		
PI	WIPI; 2002-724917/79.		
DR	P-FSDB; ABP53714.		
XX			
PT	Novel Babesia canis associated protein and nucleic acid encoding the		
PT	protein, useful in a vaccine and in the manufacture of vaccines for		
PT	combating Babesia canis infections -		
XX			
PS	Claim 20; Fig 2; 41pp; English.		
XX			
CC	The present invention describes a Babesia canis associated protein (I),		
CC	comprising a Bcvir15 protein of 15 kD molecular weight (MW) and having		
CC	a sequence of at least 80% homology to a sequence (ABP53714) of 141		
CC	amino acids, or a BcVir12 protein of 32 kD MW and having a sequence of		
CC	at least 80% homology to a sequence (ABP53715) of 285 amino acids, or		
CC	their immunogenic fragments. (I) have antiparasitic and immunostimulant		
CC	activities, and can be used in vaccines. (I) can also be used for the		
CC	preparation of a vaccine for combating B. canis infections. (I) is also		
CC	useful in a diagnostic test for the detection of antibodies against		
CC	B. canis associated antigenic material. The present sequence encodes		
CC	BcVir15 from the present invention.		
XX			
SQ	Sequence 1135 BP; 308 A; 286 C; 289 G; 252 T; 0 other;		
Alignment Scores:			
Pred. No.:	5,49e-66	Length:	1135
Score:	708.00	Matches:	141
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0
US-10-087-573-2 (1-141) x ABO82649 (1-1135)			
QY	1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu	20	
Db	75 ATGGAGTGCACATCATCACACGCCAACCCTTGTCGCCGAGAACCGTCCCACCTTTGTGTGAG	134	
QY	21 ThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMet	40	
Db	135 AC GTT TGA TGT GAT GAG GAAG CT T TGCT TCG TAAA AGT CCT CTC GAA CGC TTG GCA ATG	194	
QY	41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla	60	
Db	195 CTCAGAGCGCTGCAGGAATGTGCGGTCAACCGCTCTTCCTGGCACTTGGTCTTCTGCG	254	
QY	61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGln	80	
Db	255 ATAGCGGCAACGGTAACCCCAAAGGGGGCTTCGATGAAGCTTTAAACCAACCGCGTCCGCAG	314	
QY	81 SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr	100	
Db	315 TC AACGAAGTCTCCCGAGCTCAGGAGGCTGTCCAGGAAGATTCCGCGAAATGAATAAGACT	374	
QY	101 IleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGlu	120	
Db	375 ATAAGTCAGGAATCAGTTCGGGTAAACCAACCGGTTCGCGAAGGCCACCCCTCTCTTAGNG	434	
QY	121 LysArgAlaGluTyPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu	140	

Db 435 AAGCGGCGAGATATTTTCGTACCTTAGACTCTCTTAAGAGCCCAAGGAGTCAATAGACTC 49

Qy 141 Ile 141  
|||  
Db 495 ATC 497

RESULT 2  
ABQ82650  
ID ABQ82650 standard; cDNA; 1134 BP.  
XX AC ABQ82650;  
XX XX  
XX DT 23-DEC-2002 (first entry)  
XX DE Babesia canis BcVir32 32kD protein encoding cDNA SEQ ID NO:3.  
XX KW Babesia canis; BcVir15; 15kD protein; BcVir32; 32kD protein; infection;  
KW anti-parasitic; immunostimulant; vaccine; gene; ss.  
XX OS Babesia canis.  
XX OS  
XX FH Key Location/Qualifiers  
CD S 75..932  
FT FT /\*tag= a  
FT FT /product= "32kD protein"  
XX PN EPI238983-A1.  
XX XX  
XX PD 11-SEP-2002.  
XX PF 04-MAR-2002; 2002EP-0075830.  
XX PR 06-MAR-2001; 2001EP-0200816.  
XX PA (ALKU ) AKZO NOBEL NV.  
XX PI Schettlers TPM, Carcy BPD, Drakulovski PR, Gorenflot AP;  
DR WPI; 2002-724917/79.  
DR P-PSDB; ABP53715.  
XX XX  
PT Novel Babesia canis associated protein and nucleic acid encoding the  
PT protein, useful in a vaccine and in the manufacture of vaccines for  
PT combating Babesia canis infections -  
XX Claim 20; Fig 3; 41pp; English.

The present invention describes a Babesia canis associated protein (I), comprising a BcVir15 protein of 15 kD molecular weight (MW) and having a sequence of at least 80% homology to a sequence (ABP53714) of 141 amino acids, or a BcVir32 protein of 32 kD MW and having a sequence of at least 80% homology to a sequence (ABP53715) of 285 amino acids, or their immunogenic fragments. (I) have anti-parasitic and immunostimulant activities, and can be used in vaccines. (I) can also be used for the preparation of a vaccine for combating B. canis infections. (I) is also useful in a diagnostic test for the detection of antibodies against B. canis associated antigenic material. The present sequence encodes BcVir32 from the present invention.

SQ Sequence 1134 BP; 308 A; 286 C; 289 G; 251 T; 0 other;

Alignment Scores:  
Pred. No.: 5.32e-63 Length: 1134  
Score: 680.00 Matches: 140  
Percent Similarity: 99.29% Conservative: 0  
Best Local Similarity: 99.29% Mismatches: 1  
Query Match: 96.05% Indels: 1  
DB: 24 Gaps: 0

US-10-087-573-2 (1-141) x ABQ82650 (1-1134)

Qy 1 MetGlusThrSerThrThrThrAsnPhenValaGluAsnArgProThrPheGlyGlu 20

Db 75 ATGGAGTCGACATCAACACGACCAACTTTGTCGCGAGAACCGTCCACCTTTGGTGAG 134  
 QY 21 ThrPheAspValMetArgGluAlaLeuLeuValLysSerGluArgLeuAlaMet 40  
 Db 135 ACCTTTCATGATGATGAGGGAAGTTTCTTCGTGTAAGTCTCTGACCGCTTGGCAATG 194  
 QY 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
 Db 195 CTCAGACGCTTCAGGAATGTGGGTACCGGCTCTTCTCGCACTGTGTCTCTGCG 254  
 QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80  
 Db 255 ATAGCGGCAACGGTAACCCCAAGGGGCTTCGATGAGCTTAAACACCGCGTCGCGAG 314  
 QY 81 SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr 100  
 Db 315 TCAACGAAGTCTCGGAGCTCAGGAGCTGTCAACGGAAGATTCGCAAAATGAATAAGCT 374  
 QY 101 IleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGlu 120  
 Db 375 ATAGTCAGGAATCAGTCGGGTAAACCCCGTTGCCGGAAGCCACCTCTCTTAGAG 434  
 QY 121 LysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
 Db 435 AAGCGGCGAGAATATTTCGT-CACCTTAGATCTCTTAAAGAGCAAGAGTCAATAGACTC 493  
 QY 141 Ile 141  
 Db 494 ATC 496  
 RESULT 3  
 ABZ11244/c  
 ID ABZ11244 standard; cDNA; 2680 BP.  
 XX  
 AC ABZ11244;  
 XX  
 DT 20-JAN-2003 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 126.  
 XX  
 KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cytostatic; immunomodulator; neotropic; neuroprotective;  
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 KW haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritis; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200270539-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 05-MAR-2002; 2002WO-US05095.  
 XX  
 PR 05-MAR-2001; 2001US-0799451.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX  
 DR WPI; 2002-759812/82.  
 DR P-PSDB; ABP69027.  
 XX  
 PT New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
 PT platelet or coagulation disorders -

XX  
 PS Claim 1; SEQ ID NO 126; 1012pp + Sequence Listing; English.  
 CC  
 CC The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleoside sequence selected from any of 948 sequences  
 CC (ABZ119-ABZ1206) or their mature protein coding portion, active domain  
 CC coding protein or complementary sequences. The polynucleotides are useful  
 CC for identifying expressed genes or for physical mapping of human genome.  
 CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
 CC weight markers, as a food supplement, for generating antibodies, in  
 CC medical imaging, screening and diagnostic assays and for treating  
 CC cell-proliferative disorders (cancer), neurodegenerative diseases  
 CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
 CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
 CC disorders, platelet or coagulation disorders, wound, burns, incision,  
 CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
 CC parasitic), arthritis, etc.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2680 BP; 518 A; 863 C; 762 G; 537 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 7.73 Length: 2680  
 Score: 92.50 Matches: 38  
 Percent Similarity: 46.58% Conservative: 30  
 Best Local Similarity: 26.03% Mismatches: 52  
 Query Match: 13.06% Indels: 26  
 DB: 24 Gaps: 6  
 US-10-087-573-2 (1-141) x ABZ11244 (1-2680)  
 QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 1174 GCAGCACCCACCAACTCCCAAGGAGCTCCACGGCCCTACTAATAGC----- 1124  
 QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 Db 1123 -----AGTTGAGTTCAGTGCAGGCCCCAGTCCGCTGTGG 1088  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60  
 Db 1087 GTGATATCCAGCTCAGCAGGCACTGTGGCACTCCCGGAACCCAGGAGGCCCGTGTCT 1028  
 QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80  
 Db 1027 ATCAGCTCAGCTCAGTCCGCTGCTTCCCGCGGTCCCGGCCCGCCCGCCACGAG 968  
 QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95  
 Db 967 GTGGCTGTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 908  
 QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisArg 111  
 Db 907 TGGTTAATACCAATAGGGGAAGAGAGTGGCAGGAA---TACAGAACCCAGGCCATCG 851  
 QY 111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArgHisLe 129  
 Db 850 GCACCCCATCTCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 791  
 QY 129 uArgSerLeuLysSer 134  
 Db 790 TCGGCACTTGGAAAGT 775  
 RESULT 4  
 ABQ72569/c  
 ID ABQ72569 standard; cDNA; 4639 BP.  
 XX  
 AC ABQ72569;  
 XX  
 DT 03-SEP-2002 (first entry)  
 XX

DE Human MDDT encoding cDNA SEQ ID NO 121.

XX Human; MDDT; disease detection and treatment molecule polynucleotide;

KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;

KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;

KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;

KW hepatotropic; antiinflammatory; antiposrotic; cytostatic; anti-HIV;

KW antiallergic; antianemic; antiasthmatic; antiarteriosclerotic; antigout;

KW neuroprotective; antirheumatic; antiarthritic; gene; ss.

XX Homo sapiens.

XX WO200240715-A2.

XX 23-MAY-2002.

XX 06-SEP-2001; 2001WO-US27628.

XX 06-SEP-2000; 2000US-230505P.

XX 06-SEP-2000; 2000US-230514P.

XX 06-SEP-2000; 2000US-230515P.

XX 06-SEP-2000; 2000US-230517P.

XX 06-SEP-2000; 2000US-230518P.

XX 06-SEP-2000; 2000US-230519P.

XX 06-SEP-2000; 2000US-230595P.

XX 06-SEP-2000; 2000US-230597P.

XX 06-SEP-2000; 2000US-230598P.

XX 06-SEP-2000; 2000US-230599P.

XX 06-SEP-2000; 2000US-230610P.

XX 06-SEP-2000; 2000US-230865P.

XX 06-SEP-2000; 2000US-230988P.

XX 06-SEP-2000; 2000US-230989P.

XX 07-SEP-2000; 2000US-230951P.

XX 07-SEP-2000; 2000US-231163P.

XX 07-SEP-2000; 2000US-231167P.

XX (INCY-) INCYTE GENOMICS INC.

XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;

PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;

PI Momiya MG, Bradley DL, Rohatgi SD, Harris B, Rosberry AM;

PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;

PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;

XX WPI; 2002-527544/56.

DR P-PSDB; ABP51351.

XX Novel human disease detection and treatment polypeptide, useful in

PT diagnosis, prevention or treatment of cell proliferative disorders e.g.

PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder

PT e.g. AIDS -

XX Claim 1; Page 375-376; 618pp; English.

XX The invention relates to an isolated human disease detection and

XX treatment (MDDT) polypeptide (I) selected from a polypeptide having a

XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the

XX specification, a naturally occurring polypeptide comprising a sequence

XX having at least 90% identity to (I) or a biologically active or

XX immunogenic fragment of (I). (I) is useful for screening a compound for

XX effectiveness as an agonist or antagonist, for screening a compound that

XX specifically binds (I) or modulates the activity of (I), and for

XX preparing a polyclonal or monoclonal antibody by hybridoma technology.

XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for

XX screening a compound for effectiveness in altering expression of a target

XX polynucleotide comprising. Oligonucleotides and antibodies are useful for

XX detecting MDDT in a sample or for assessing toxicity of a test compound,

XX in a diagnostic test for a condition or a disease associated with the

XX expression of MDDT in a biological sample, for detecting (I) in a sample,

XX and for purifying (I) from a sample. A composition comprising (I), an

XX agonist or antagonist is useful for treating a disease or condition

XX associated with decreased or increased expression of functional MDDT.

XX (I) or (II) are useful for diagnosing, treating or preventing disorders

CC associated with aberrant expression of MDDT, where the disorders are

CC selected from a cell proliferative disorder such as arteriosclerosis,

CC cirrhosis, hepatitis, psoriasis, and cancer and an

CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,

CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or

CC rheumatoid arthritis. (II) are useful for creating knockin humanised

CC animals or transgenic animals to model human diseases, in somatic or

CC germline gene therapy, to generate a transcript image of a tissue or cell

CC type, for detecting differences in the chromosomal location due to

CC translocation or inversion among normal, carrier or affected individuals

CC and as hybridisation probes for mapping naturally occurring genomic

CC sequences.

XX SQ Sequence 4839 BP; 946 A; 1467 C; 1418 G; 1008 T; 0 other;

XX

Alignment Scores:

Pred. No.: 16.8 Length: 4839

Score: 92.50 Matches: 38

Percent Similarity: 46.58% Conservativity: 30

Best Local Similarity: 26.03% Mismatches: 52

Query Match: 13.06% Indels: 26

DB: 24 Gaps: 6

US-10-087-573-2 (1-141) x ABQ72569 (1-4839)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22

DB 2839 GCAGCACACACACACACACACACACAGGAGGCTCCAGCGCCCTACTAATAGC----- 2789

QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42

DB 2788 -----AGGTTGAGGTGACGTGCGCCAGCCGCGCGTGTGG 2753

QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60

DB 2752 GTGATATCCAGGTCACAGGCACTGTGGCACTCCGGGAACACAGGAGGCGCCCGTCT 2693

QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80

DB 2692 ATCATCTCAGCTCAGCTCCGCTGCTTCCCGCGGGTCCGCGCCCGCCACCCAGAG 2633

QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95

DB 2632 GTGGCTGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2573

QY 96 -----GluMetAsnLysThrLysSerGlnGluSerAlaArgVal-AsnHisArg 111

DB 2572 TGGTTAATACCAATAGGGGAGAGAGTGGCAAGGAA---TACAGAACAGGCCATCG 2516

QY 111 gluProGluGlyHis-----ProLeuLeuLysArgAlaGluTyrPheArgHisLe 129

DB 2515 GCACCCCATCTCATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2456

QY 129 uArgSerLeuLysSer 134

DB 2455 TCGGCACTTGGAGT 2440

RESULT 5

ABK15176/c

ID ABK15176 standard; DNA; 10123 BP.

XX AC ABK15176;

XX 23-APR-2002 (first entry)

XX Human REPTR 8 cDNA sequence.

XX REPTR; human; antiinflammatory; cytostatic; immunosuppressive;

XX antiviral; anti-HIV; antiarthritic; anticonvulsant; nootropic;

XX neuroprotective; antiallergic; antibody; immunogen; endometriosis;

XX gastrointestinal disorder; gastritis; oesophageal carcinoma;

XX Crohn's disease; irritable bowel syndrome; ulcerative colitis;

XX endocrine disorder; hypothyroidism; Kallman's disease;







XX The invention relates to a novel method for identifying a candidate p53  
 CC pathway modulating agent. The method comprises providing an assay system  
 CC comprising a purified cadherin (CAD) polypeptide or nucleic acid, or  
 CC their functionally active fragment or derivative. The method is useful  
 CC for identifying a candidate p53 pathway modulating agent, modulating a  
 CC p53 pathway of a cell and for diagnosing a disease in a patient. In  
 CC particular, the disease is cancer, e.g. breast cancer, colon cancer,  
 CC kidney cancer, lung cancer or cancer of the ovary, which has an  
 CC expression level of greater than 25%. The identified modulators are  
 CC useful as targets for novel therapeutics. This polynucleotide sequence  
 CC represents a cadherin (CAD) nucleic acid of the invention.

XX Sequence 10531 BP; 1980 A; 3340 C; 3141 G; 2070 T; 0 other;

Alignment Scores:  
 Pred. No.: 46.8 Length: 10531  
 Score: 92.50 Matches: 38  
 Percent Similarity: 46.58% Conservative: 30  
 Best Local Similarity: 26.03% Mismatches: 52  
 Query Match: 13.06% Indels: 26  
 DB: 25 Gaps: 6

US-10-087-573-2 (1-141) x AAL54474 (1-10531)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 9037 GCAAGCACCACTCCCAAGGAGGCTGCCAGGCCCTCTACTAATAGC----- 8987  
 QY 23 AspValMetArgGluAlaLeuLeuValLysSerSerGluArgLeuAlaMetLeuArg 42  
 Db 8986 -----AGGTTGAGGTTCAGGTGCCAGGCCCAAGTCCGGTGTGG 8951  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60  
 Db 8950 GTGATATCCACGGTCACAGGCACTTGGCACTCCGGAAACAGGAGGCCCGCTGCT 8891  
 QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80  
 Db 8890 ATCACCTCCAGCTCAGCTCCGCTGTGCTTCCCGCGGGTCCGCGCCCAACCCACAG 8831  
 QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95  
 Db 8830 GTGGCTGTTCCGCTGCTGCTGCCAGCTGTCCACCGCAGGTACAGGCTCCTGTAGTC 8771  
 QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisArg 111  
 Db 8770 TGGTTAATACCAAAATAGGGGAAGAGGTGGCAAGGAA---TACAGAACAGGCAATCG 8714  
 QY 111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluThrPheArgHisLe 129  
 Db 8713 GCACCCCACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8654  
 QY 129 uArgSerLeuLysSer 134  
 Db 8653 TCGGGCACTTGAAGT 8638

RESULT 7  
 ID ABA93093/c  
 ID ABA93093 standard; cDNA; 10759 BP.

AC ABA93093;  
 XX 12-APR-2002 (first entry)  
 DT Human dachshous encoding cDNA SEQ ID NO:1.

DE Dachshous; Freeman-Sheldon syndrome; congenital anamorphosis syndrome;  
 XX human; gene; ss.  
 KW Homo sapiens.  
 XX Key Location/Qualifiers

FT CDS 411..10307  
 FT /\*tag= a  
 FT /product= "dachshous protein"  
 PN JP2001327295-A.  
 XX 27-NOV-2001.  
 XX 27-JUL-2000; 2000JP-0226291.  
 XX 14-MAR-2000; 2000JP-0069599.  
 XX (KAZU-) 2H KAZUSA DNA KENKYUSHO.  
 XX WP1; 2002-127071/17.  
 DR P-PSDB; ABB05430.  
 XX A human dachshous gene and a protein encoded by the gene -  
 PT Claim 2; Page 10-27; 40pp; Japanese.  
 XX The present sequence encodes a human dachshous protein. The dachshous  
 CC gene can be used for diagnosing or treating Freeman-Sheldon syndrome  
 CC or congenital anamorphosis syndrome.  
 XX SQ Sequence 10759 BP; 2008 A; 3412 C; 3256 G; 2083 T; 0 other;

Alignment Scores:

Pred. No.: 48.1 Length: 10759  
 Score: 92.50 Matches: 38  
 Percent Similarity: 46.58% Conservative: 30  
 Best Local Similarity: 26.03% Mismatches: 52  
 Query Match: 13.06% Indels: 26  
 DB: 24 Gaps: 6

US-10-087-573-2 (1-141) x ABA93093 (1-10759)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 9265 GCAAGCACCACTCCCAAGGAGGCTGCCAGGCCCTCTACTAATAGC----- 9215  
 QY 23 AspValMetArgGluAlaLeuLeuValLysSerSerGluArgLeuAlaMetLeuArg 42  
 Db 9214 -----AGGTTGAGGTTCAGGTGCCAGGCCCAAGTCCGGTGTGG 9179  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeu-----ProGlyThrGlyAlaSerAla 60  
 Db 9178 GTGATATCCACGGTCACAGGCACTTGGCACTCCGGAAACAGGAGGCCCGCTGCT 9119  
 QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArgProGln 80  
 Db 9118 ATCACCTCCAGCTCAGCTCCGCTGTGCTTCCCGCGGGTCCGCGCCCAACCCACAG 9059  
 QY 81 ---SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg----- 95  
 Db 9058 GTGGCTGTTCCGCTGCTGCTGCCAGCTGTCCACCGCAGGTACAGGCTCCTGTAGTC 8999  
 QY 96 -----GluMetAsnLysThrIleSerGlnGluSerAlaArgVal-AsnHisArg 111  
 Db 8998 TGGTTAATACCAAAATAGGGGAAGAGGTGGCAAGGAA---TACAGAACAGGCAATCG 8942  
 QY 111 gLeuProGluGlyHis-----ProLeuLeuGluLysArgAlaGluThrPheArgHisLe 129  
 Db 8941 GCACCCCACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8882  
 QY 129 uArgSerLeuLysSer 134  
 Db 8881 TCGGGCACTTGAAGT 8866

RESULT 8  
 ID ABA92651  
 ID ABA92651 standard; DNA; 90 BP.

AC ABQ82651;  
 XX  
 DT 23-DEC-2002 (first entry)  
 XX  
 DE Babesia canis Bcivr oligonucleotide PCR primer SEQ ID NO:5.  
 XX  
 KW Babesia canis; Bcivr15; 15kD protein; Bcivr32; 32kD protein; infection;  
 XX antiparasitic; immunostimulant; vaccine; PCR primer; ss.  
 XX  
 OS Babesia canis.  
 XX  
 PN EP1238983-A1.  
 XX  
 PD 11-SEP-2002.  
 XX  
 PF 04-MAR-2002; 2002EP-0075830.  
 XX  
 PR 06-MAR-2001; 2001EP-0200816.  
 XX  
 PA (ALKU ) AKZO NOBEL NV.  
 XX  
 PI Schettlers TPM, Carcy BPD, Drakulovski PR, Gorenflot AF;  
 XX  
 DR WPI; 2002-724917/79.  
 XX  
 PT Novel Babesia canis associated protein and nucleic acid encoding the  
 PT protein, useful in a vaccine and in the manufacture of vaccines for  
 PT combating Babesia canis infections. -  
 XX  
 PS Example 2; Page 12; 41pp; English.  
 XX  
 CC The present invention describes a Babesia canis associated protein (I),  
 CC comprising a Bcivr15 protein of 15 kD molecular weight (MW) and having  
 CC a sequence of at least 80% homology to a sequence (ABP53714) of 141  
 CC amino acids, or a Bcivr32 protein of 32 kD MW and having a sequence of  
 CC at least 80% homology to a sequence (ABP53715) of 285 amino acids, or  
 CC their immunogenic fragments. (I) have antiparasitic and immunostimulant  
 CC activities, and can be used in vaccines. (I) can also be used for the  
 CC preparation of a vaccine for combating B. canis infections. (I) is also  
 CC useful in a diagnostic test for the detection of antibodies against  
 CC B. canis associated antigenic material. The present invention represents  
 CC a PCR primer for Babesia canis Bcivr which is used in an example from  
 CC the present invention.  
 XX  
 SQ Sequence 90 BP; 26 A; 27 C; 19 G; 18 T; 0 other;

KW nutritional condition; peripheral nervous system disorder; ischaemia;  
 KW diabetic neuropathy; autoimmune haemolytic anaemia; respiratory system;  
 KW oral cavity; gastrointestinal tract; liver; cirrhosis; pancreatitis;  
 KW polycystic renal disease; urinary tract; genitalia; endometriosis;  
 KW breast; chronic mastitis; thyroid gland; Hashimoto's thyroiditis;  
 KW adrenal gland; skin; psoriasis; muscular atrophy; bone marrow;  
 XX osteoporosis; cancer; autoimmune disease; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200176532-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 09-APR-2001; 2001WO-US11655.  
 XX  
 PR 11-APR-2000; 2000US-0547596.  
 XX  
 PA (COGR-) COGENT NEUROSCIENCE INC.  
 XX  
 PI Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC, Lo DC;  
 XX  
 DR WPI; 2002-017408/02.  
 XX  
 DR P-PSDB; AAU69691.  
 XX  
 PT Novel nucleic acids referred as protective sequences and their encoded  
 PT products for diagnosing, treating diseases involving cell death,  
 PT including neurological disorders e.g. stroke and for identifying  
 PT modulators -  
 XX  
 PS Claim 2; Figure 10A; 256pp; English.  
 XX  
 CC The invention relates to isolated protective sequence polypeptides (I)  
 CC and polynucleotides (II). (I) is useful for transferring a protective  
 CC sequence into a cell, which delays and/or prevents the cell from  
 CC undergoing cell death. Protective sequences, their products or  
 CC antibodies are useful diagnostically, prophylactically, therapeutically  
 CC or as targets for treatment and diagnosis of conditions, disorders or  
 CC diseases involving cell death. The protective sequences and their  
 CC products are useful for preventing or treating disorders of the central  
 CC nervous system including neurological and psychiatric conditions,  
 CC cerebral oedema, infections such as meningitis, degenerative diseases  
 CC such as Alzheimer's and motor neuron disease, demyelinating diseases such  
 CC as multiple sclerosis, nutritional conditions, disorders of the  
 CC peripheral nervous system including diabetic neuropathy, disorders  
 CC which cause cell death in organ systems including blood vessels, heart  
 CC (ischaemia), blood cells (autoimmune haemolytic anaemia), respiratory  
 CC system (asthma), oral cavity, gastrointestinal tract, liver (cirrhosis),  
 CC pancreatitis, polycystic renal disease, urinary tract, genitalia  
 CC (congenital anomalies), endometriosis, breast (chronic mastitis), thyroid  
 CC gland (Hashimoto's thyroiditis), adrenal gland, skin (psoriasis),  
 CC musculoskeletal system (muscular atrophy) bone marrow or bone  
 CC (osteoporosis). The compositions promote cell death and are useful for  
 CC treating and/or ameliorating cancer and autoimmune diseases. The  
 CC compounds are further useful for treating physiological impacts on  
 CC organs caused by infection which induce cell death. (I) is useful to  
 CC raise an immune response, as a reagent in assays designed to  
 CC quantitatively determine levels of the protein in biological fluids, as  
 CC markers for tissues in which the corresponding protein is expressed and  
 CC to isolate receptors or ligands. AAS62967-AAS63142 represent the  
 CC protective polynucleotide sequences as described in the invention.  
 XX  
 SQ Sequence 1386 BP; 268 A; 416 C; 460 G; 242 T; 0 other;

Alignment Scores:  
 Pred. No.: 6 Length: 1386  
 Score: 90.00 Matches: 40  
 Percent Similarity: 41.54% Conservative: 14  
 Best Local Similarity: 30.77% Mismatches: 41  
 Query Match: 12.71% Indels: 35  
 DB: 24 Gaps: 6  
 US-10-087-573-2 (1-141) x AAS63094 (1-1386)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPhe 18  
 DB 35 ATGGAGTCGACATCAACACGACCAACTTGTGGCGAGACCGCCACCTTT 88  
 RESULT 9  
 AAS63094/C  
 ID AAS63094 standard; cDNA; 1386 BP.  
 XX  
 AC AAS63094;  
 XX  
 DT 29-JAN-2002 (first entry)  
 XX  
 DE Cell death protective sequence CNI-00729, ORF #1.  
 XX  
 KW Human; protective sequence; cell death; cerebral oedema; infection;  
 KW meningitis; degenerative disease; Alzheimer's disease; heart disease;  
 KW motor neuron disease; demyelinating disease; multiple sclerosis; asthma;

CC cell movement.

XX Sequence 1657 BP; 359 A; 497 C; 511 G; 290 T; 0 other;

SQ

Alignment Scores:

Pred. No.: 7.58 Length: 1657

Score: 90.00 Matches: 40

Percent Similarity: 41.54% Conservative: 14

Best Local Similarity: 30.77% Mismatches: 41

Query Match: 12.71% Indels: 35

DB: 17 Gaps: 6

US-10-087-573-2 (1-141) x AAT30360 (1-1657)

QY 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26

DB 500 ACACGCTGGTGTGGCCCTCCAGGGTGACGAGGGCTCCCGAGGGCGACATCAGGCC 441

QY 27 -----GluAlaLeuLeuArgVallySerSerGluArgLeuAlaMetLeuArg 42

DB 440 CCATCCGGGATCTCCACACCATGATCTGTCAGTCTCGAGGCCACTGGCAATGACGTTG 381

QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

DB 380 TCATTTG-----TGGGGCCACCGGCGATGCTAGCACAGGGGCTGTGTGGCCACAG 330

QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74

DB 329 ACCGTGGCGCATCTTTGTCACACAGTCCAGTCTTGGCCAGGGGCGACACAGGAAGGCC 270

QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLysle 94

DB 269 CTCTCCCG-----CTGGGCTCACAGATC 246

QY 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108

DB 245 AGGGCCACAACTTAGGTTGACAGACAGAACCCACTGTCCAGGTGGTCTGTGAGACG 186

QY 108 IAsnHisArgLeuProGluGlyHisPro 117

DB 185 CGCACATCTTCATAGCACTGGTCCGCT 158

RESULT 11

AAK54941/c

ID AAK54941 standard; cDNA; 1659 BP.

XX

AC AAK54941;

XX

DT 13-NOV-2001 (first entry)

XX

DE Human haematological malignancy-related antigen coding sequence #666.

XX

KW Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;

KW haematological malignancy; antigen; chronic lymphocytic leukaemia;

XX follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma; ss.

XX

OS Homo sapiens.

XX

PN WO200164886-A2.

XX

PD 07-SEP-2001.

XX

PF 01-MAR-2001; 2001WO-US07272.

XX

PR 01-MAR-2000; 2000US-0186126.

PR 17-MAR-2000; 2000US-0190479.

PR 27-APR-2000; 2000US-0200545.

PR 28-APR-2000; 2000US-0200303.

PR 28-APR-2000; 2000US-0200779.

PR 01-MAY-2000; 2000US-0200999.

PR 04-MAY-2000; 2000US-0202084.

PR 22-MAY-2000; 2000US-0206201.

PR 14-JUL-2000; 2000US-0218950.

QY 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26

DB 401 ACACGCTGGTGTGGCCCTCCAGGGTGACGAGGGCTCCCGAGGGCGACATCAGGCC 342

QY 27 -----GluAlaLeuLeuArgVallySerSerGluArgLeuAlaMetLeuArg 42

DB 341 CCATCTGGGATCTCCACACCATGCTGTCAGTCTCGAGGCCACTGGCAATGACGTTG 282

QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

DB 281 TCATTTG-----TGGGGCCACCGGCGATGCTAGCACAGGGGCTGTGTGGCCACAG 231

QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74

DB 230 ACCGTGGCGCATCTTTGTCACACAGTCCAGTCTTGGCCAGGGGCGACACAGGAAGGCC 171

QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLysle 94

DB 170 CCTCCCGCG-----CTGGGCTCACAGATC 147

QY 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108

DB 146 AGGGCCACAACTTAGGTTGACAGACAGAACCCACTGTCCAGGTGGTCTGTGAGACG 87

QY 108 IAsnHisArgLeuProGluGlyHisPro 117

DB 86 CGCACATCTTCATAGCACTGGTCCGCT 59

RESULT 10

AAT30360/c

ID AAT30360 standard; cDNA; 1657 BP.

XX

AC AAT30360;

XX

DT 26-NOV-1996 (first entry)

XX

DE Human p57 coding sequence.

XX

KW WD40 repeated structure region; human p57 protein; dimer;

KW anti-cancer; anti-HIV agent; actin; cell movement; ds.

XX

OS Homo sapiens.

XX

XX Key Location/Qualifiers

FF 100..1485

FT CDS

FT /\*tag= a

FT /product= p57

XX

XX JF08119996-A.

XX

PD 14-MAY-1996.

XX

PF 21-OCT-1994; 94JP-0282743.

XX

PR 21-OCT-1994; 94JP-0282743.

XX

XX (NISB ) JAPAN TOBACCO INC.

XX

XX WPI; 1996-283507/29.

DR P-PSDB; AAR98341.

XX

PT A new protein, p57, comprising WD40 repeat region - used for

PT development of anti-cancer and anti-HIV agents

XX

XX Claim 9; Page 23-25; 51pp; Japanese.

XX

CC This sequence encodes the human p57 protein. p57 contains a leucine

CC rich C-terminal peptide which comprises a leucine residue after

CC each 7 amino acids (see also R983340) and a WD40 repeated structure

CC region (see also AAR98334-38) containing five WD40 regions. p57 forms a

CC dimer. It can be used in the development of an anti-cancer agent and

CC an anti-HIV agent. It can also combine with actin in the control of

PR 03-AUG-2000; 2000US-0222903.  
PR 04-AUG-2000; 2000US-0223416.  
PR 07-AUG-2000; 2000US-0223378.  
XX (CORI-) CORIXA CORP.  
XX Gaiger A, Algate PA, Mannion J;  
XX WPI; 2001-514842/56.  
XX Compositions and methods for the detection of hematological  
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular  
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -  
XX Claim 31; Pages 504-505; 1252pp; English.  
XX The present invention relates to compositions and methods for the  
CC detection, diagnosis and therapy of haematological malignancies. The  
CC present sequence is the coding sequence of a human haematological  
CC malignancy related antigen. The methods of the present invention comprise  
CC detecting the presence of haematological malignancy related antigen(s) in  
CC a sample obtained from the patient (an increased level of the  
CC polypeptide, compared to an unaffected individual, is indicative of an  
CC increased risk). Haematological malignancies which can be treated using  
CC the present invention are chronic lymphocytic leukaemia, lymphoma,  
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B  
CC cell non-Hodgkin's lymphoma.  
XX Sequence 1659 BP; 360 A; 497 C; 511 G; 291 T; 0 other;  
Alignment Scores:  
Pred. No.: 7.6 Length: 1659  
Score: 90.00 Matches: 40  
Percent Similarity: 41.54% Conservative: 14  
Best Local Similarity: 30.77% Mismatches: 41  
Query Match: 12.71% Indels: 35  
DB: 22 Gaps: 6  
US-10-087-573-2 (1-141) x AAK54941 (1-1659)  
Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg --- 26  
Db 501 ACACGCTGGTGGCCCTCCAGGTGACGACGGCTCCCGAGGGCAGCATCAGGCC 442  
Qy 27 -----GluAlaLeuLeuArgVallySerSerGluArgLeuAlaMetLeuArg 42  
Db 441 CCATCCGGGATCTCCACACCATGACTGTGCAGTCTCTGGAGCCACTGGCAATGACGTTG 382  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla ----- 58  
Db 381 TCAITG -----TGGGGCCACAGCGCATGTCTAGCACAGGGGTGTGTGGCCACAG 331  
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 330 ACCGTGGGCGCATCTTGTCCACACGTCACGTCCTGCCAGGGSCACACGGAAGGCC 271  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLeu 94  
Db 270 CCTCCCGCG -----CTGGGCTCACAGATC 247  
Qy 95 ArgGluMetAsn -----LysThrIleSerGlnGlu-SerAlaArgVa 108  
Db 246 AGGCCCAAACTTAGGTTGCACAGCAGAGAAGCCACTGTCCAGGGTGTGTGAGACG 187  
Qy 108 LAsnHisArgLeuProGluGlyHisPro 117  
Db 186 CGCACATCTTCATAGCACTGGTCGGCCT 159  
RESULT 12  
ABK83666/c  
ID ABK83666 standard; cDNA; 1659 BP.  
XX  
AC ABK83666;

XX  
DT 14-AUG-2002 (first entry)  
XX Human cDNA differentially expressed in granulocytic cells #237.  
DE  
XX  
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX  
OS Homo sapiens.  
XX  
PN WO200228999-A2.  
PD 11-APR-2002.  
XX  
PF 03-OCT-2001; 2001WO-US30821.  
PR 03-OCT-2000; 2000US-237189P.  
XX (GENE-) GENE LOGIC INC.  
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
XX WPI; 2002-435328/46.  
DR  
XX  
PT Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
PT drug toxicity -  
XX  
PS Claim 1; SEQ ID No 237; 114pp; English.  
XX  
CC The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated  
CC GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal  
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion syndrome,  
CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
CC periodontal disease; also bacterial infection, viral infection,  
CC parasitic infection, protozoal infection, fungal infection and M5 is  
CC useful for treating one of the above conditions. The present  
CC sequence represents a gene differentially expressed in granulocytes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

WPI: 2002-017408/02.

antibodies are useful diagnostically, prophylactically, therapeutically or as targets for treatment and diagnosis of conditions, disorders or diseases involving cell death. The protective sequences and their products are useful for preventing or treating disorders of the central nervous system including neurological and psychiatric conditions,

cerebral oedema, infections such as meningitis, degenerative diseases such as Alzheimer's and motor neuron disease, demyelinating diseases such

peripheral nervous system including diabetic neuropathy, disorders which cause cell death in organ systems including blood vessels, heart

system (asthma), oral cavity, gastrointestinal tract, liver (cirrhosis)  
pancreatitis, polycystic renal disease, urinary tract, genitalia

gland (Hashimoto's thyroiditis), adrenal gland, skin (psoriasis), musculoskeletal system (muscular atrophy), bone marrow or bone (osteoporosis). The concentration of monocyte cell death and are useful for

compounds are further useful for treating physiological impacts on treating and/or ameliorating cancer and autoimmune diseases. The compounds are further useful for inducing cell death. (I) is useful to

quantitatively determine levels of the protein in biological fluids, as markers for tissues in which the corresponding protein is expressed and raise an immune response, as a reagent in assays designed to

protective polynucleotide sequences as described in the invention.

ment Scores:

```

e:
ent Similarity:
90.00
41.54%

```

DE Cell death protective sequence CNI-00729.  
XX

0-087-573-2 (1-141) X AAS63093 (1-1659)

540 ACACGCTTGGTGGCCCTCCAGGGTGACGACGGGCTCCGCAGGGGCAGCATCAGGCC 4

27 -----GluAraDeuWaiGValDpsISeiSeiGmniGDeuWaiInUeCaaS  
:::  
480 CCATCTGGGATCTCCCAACCATAGACTGTGCAGTCTCGGAGGCCACTGGCAATGACGTTG

43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla

59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 7

369 ACCGTGGGGCGCATCTTGTGCCACACGTCAGTCTTGTCCAGGGGCGAGCACCGAAGGCC

309 CCTCCCCG-----CTGGCTCACAGATC 2

QY 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108  
Db 285 AGGCCCAAACTAGGTTGACAGACAGAACCCACTGCCAGGTGCTGTGAGAGC 226  
QY 108 lAsnHisArgLeuProGluGlyHisPro 117  
Db 225 CGCACATCTTCATAGCACTGGTCGGCCT 198  
RESULT 14  
AAT30370/c  
ID AAT30370 standard; cDNA; 1842 BP.  
XX AC AAT30370;  
XX 26-NOV-1996 (first entry)  
XX GST/truncated human p57 fusion protein coding sequence.  
XX WD40 repeated structure region; bovine p57 protein; dimer;  
XX anti-cancer; anti-HIV agent; actin; cell movement; ds.  
XX Synthetic.  
XX JP08119996-A.  
XX 14-MAY-1996.  
XX 21-OCT-1994; 94JP-0282743.  
XX 21-OCT-1994; 94JP-0282743.  
XX (NISB ) JAPAN TOBACCO INC.  
XX WPI; 1996-283507/29.  
XX P-PSDB; AAR98344.  
XX A new protein, p57, comprising WD40 repeat region - used for  
XX development of anti-cancer and anti-HIV agents  
XX Disclosure; Page 33-35; 51pp; Japanese.  
XX This sequence encodes a glutathione S transferase/truncated human p57  
XX fusion protein. This sequence is derived from a derivative of the  
XX plasmid pGEX/hp57. This plasmid is used in the expression of the  
XX human p57 protein. p57 contains a leucine rich C-terminal peptide  
XX which comprises a leucine residue after each 7 amino acids and a  
XX WD40 repeated structure region containing five WD40 regions. p57  
XX forms a dimer. It can be used in the development of an anti-cancer  
XX agent and an anti-HIV agent. It can also combine with actin in the  
XX control of cell movement.  
XX Sequence 1842 BP; 422 A; 462 C; 532 G; 426 T; 0 other;  
SQ Alignment Scores:  
Pred. No.: 8.72 Length: 1842  
Score: 90.00 Matches: 40  
Percent Similarity: 41.54% Conservative: 14  
Best Local Similarity: 30.77% Mismatches: 41  
Query Match: 12.71% Indels: 35  
DB: 17 Gaps: 6  
US-10-087-573-2 (1-141) x AAT30370 (1-1842)  
QY 8 ThrAsnPhaValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26  
Db 1085 ACACGGTTGGTGGCCCTCCAGGTTGACAGCGGCTCCCGAGGGGCGCATCAGGCC 1026  
QY 27 -----GluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 1025 CCATCGGGATCTCCACACCCAGTGTGAGTCTCCGAGCCACTGGCAATGACGTTG 966  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 965 TCATTG-----TGGGGGACACGCGGATGCTAGCACAGGGGCTGTGGCCACAG 915  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 914 ACCGTGGCGCATCTTGTCCACACGTCAGTCTTGGCCAGGGGCGACCCAGGAGGCC 855  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 854 CCTCCCGC-----CTGGCTTCACAGATC 831  
QY 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108  
Db 830 AGGGCCCAAACTTAGGGTTGACAGACAGAACCCACTGTCCAGGTGCTGTGAGAGC 771  
QY 108 lAsnHisArgLeuProGluGlyHisPro 117  
Db 770 CGCACATCTTCATAGCACTGGTCGGCCT 743  
RESULT 15  
AAT30369/c  
ID AAT30369 standard; cDNA; 2067 BP.  
XX AC AAT30369;  
XX 26-NOV-1996 (first entry)  
XX GST/human p57 fusion protein coding sequence.  
XX WD40 repeated structure region; bovine p57 protein; dimer;  
XX anti-cancer; anti-HIV agent; actin; cell movement; ds.  
XX Synthetic.  
XX JP08119996-A.  
XX 14-MAY-1996.  
XX 21-OCT-1994; 94JP-0282743.  
XX 21-OCT-1994; 94JP-0282743.  
XX (NISB ) JAPAN TOBACCO INC.  
XX WPI; 1996-283507/29.  
XX P-PSDB; AAR98343.  
XX A new protein, p57, comprising WD40 repeat region - used for  
XX development of anti-cancer and anti-HIV agents  
XX Example 5; Page 30-32; 51pp; Japanese.  
XX This sequence encodes a glutathione S transferase/human p57 fusion  
XX protein. This sequence is derived from the plasmid pGEX/hp57. This  
XX plasmid is used in the expression of the human p57 protein. p57  
XX contains a leucine rich C-terminal peptide which comprises a leucine  
XX residue after each 7 amino acids and a WD40 repeated structure region  
XX containing five WD40 regions. p57 forms a dimer. It can be used  
XX in the development of an anti-cancer agent and an anti-HIV agent.  
XX It can also combine with actin in the control of cell movement.  
XX Sequence 2067 BP; 470 A; 532 C; 612 G; 453 T; 0 other;  
SQ Alignment Scores:  
Pred. No.: 10.1 Length: 2067  
Score: 90.00 Matches: 40  
Percent Similarity: 41.54% Conservative: 14  
Best Local Similarity: 30.77% Mismatches: 41  
Query Match: 12.71% Indels: 35  
DB: 17 Gaps: 6  
US-10-087-573-2 (1-141) x AAT30369 (1-2067)

QY 8 ThrAsnPhaValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26  
DB 1085 ACACGGTTGGTGGCCCTCCAGGGTGACAGGGCTCCCGAGGGGACATCAGGCC 1026  
QY 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
DB 1025 CCATCCGGGATCTCCACACCATGACTGTGCAGTCTCCGAGCCCACTGGCAATGACGTTG 966  
QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
DB 965 TCATTG-----TGGCGGACACAGGGCGTGTAGCACAGGGGCTGTGGCCACAG 915  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
DB 914 ACGTGGGGCGCATCTGTGTCACACGCTCCAGTCTTGCCCGAGGGCAGCACAGGAAGGCC 855  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
DB 854 CTTCCCGCG-----CTGCCCTCACAGATC 831  
QY 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108  
DB 830 AGGGCCACAACTTAGGGTTGACAGCACAGAACGACACTGTCCAGGTGTGTGTGAGACG 771  
QY 108 lAsnHisArgLeuProGluGlyHisPro 117  
DB 770 CGCACATCTTCATAGCACTGGTGGCCT 743  
RESULT 16  
ABQ46392/c  
ID ABQ46392 standard; DNA; 1007 BP.  
XX AC ABQ46392;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 32983.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EPI0074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ3410-ABQ411 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 1007 BP; 117 A; 140 C; 381 G; 369 T; 0 other;  
Alignment Scores:  
Pred. No.: 8.23 Length: 1007  
Score: 87.00 Matches: 29  
Percent Similarity: 47.06% Conservative: 11  
Best Local Similarity: 34.12% Mismatches: 29  
Query Match: 12.29% Indels: 16  
DB: 24 Gaps: 5  
US-10-087-573-2 (1-141) x ABQ46392 (1-1007)  
QY 54 ProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLys 73  
DB 392 CCGTCAACCCCAACACACCATCTAGCGCGCAACCCCTACCCCGCGCTCCCGCAACCCC 333  
QY 74 Leu---LysProProArgProGln-----SerThrLysSerProGluLeuArg--- 88  
DB 332 CTTCCGAGACCGCCCGCGCTCCCTAAACCCCGAATTAAATATACCCGAAATTCGACGC 273  
QY 89 ---GluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArg 107  
DB 272 TACGAACTCGATACGAAAAACGCGACGAAAAAATTTTACA-----AAAACCGCAAC 222  
QY 108 ValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheArg 127  
DB 221 GAAATATCACCGTCGACCG-----CGCCAAAAACAAATCTATCAC 183  
QY 128 HisLeuArgSerLeu 132  
DB 182 CGTCTCCGACCCCTA 168  
RESULT 17  
ABQ46393  
ID ABQ46393 standard; DNA; 1007 BP.  
XX AC ABQ46393;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 32984.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EPI0074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.



[illegible]

Qy 43 AlaLeu-----AlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59  
 Db 214439 GGATTGTGTATCGCGGCTTCTGCTCT-----CCAGGCGGGAGTGACC 214392  
 Qy 60 AlaIleAlaIaThrValThrProLysGlyAla-----SerMet 72  
 Db 214391 TCGTGCGCTGACCGTGACTTCGAACGTCGCTTCAGCCAGCATAAAGAGCGGTGATG 214332  
 Qy 73 LysLeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArg 88  
 Db 214331 GTCTCTGCTGCTGCGGTGGCAGAAATCAAGCCAGTCAGATCAGATCAGTTGGA 214272  
 Qy 89 GluLeuSerArgLysIleArgGluMetAsnLysThrIle-----SerGlnGlu 104  
 Db 214271 TCAATCAGCTTGGCGAGCGTTTCAGTATCCCGACATTTGTAATGCGGATACCTTCACA 214212  
 Qy 105 SerAlaArgValAsn-HisArgLeuProGluGlyHisPro 117  
 Db 214211 TCGGCAGCGTTGAACCTATCGCTAGCTGATAGGAATCCT 214172  
 RESULT 19  
 AAF13411  
 ID AAF13411 standard; cDNA; 1337 BP.  
 AC AAF13411;  
 DT 13-MAR-2001 (first entry)  
 DE Aspergillus oryzae EST SEQ ID NO:5934.  
 KW Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 OS Aspergillus oryzae.  
 XX  
 XX WO200056762-A2.  
 XX  
 XX 28-SEP-2000.  
 XX  
 XX 22-MAR-2000; 2000WO-US07781.  
 XX  
 XX 22-MAR-1999; 99US-0273623.  
 XX  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 XX (NOVO ) NOVO NORDISK AS.  
 XX  
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 WPI; 2000-594572/56.  
 XX  
 XX Monitoring differential expression of genes in filamentous fungal cells  
 XX uses fluorescence-labeled nucleic acids isolated from the cells and a  
 XX substrate of expressed sequence tags -  
 XX  
 XX Claim 88; Page 2451; 3161pp; English.  
 XX  
 XX The present invention describes a method for monitoring differential  
 XX expression of genes in a first filamentous fungal (FF) cell relative to  
 XX expression of the same genes in one or more second filamentous fungal  
 XX cells. The method uses fluorescence-labeled nucleic acids isolated from  
 XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 XX are used in the methods for monitoring differential expression of genes  
 XX in a first filamentous fungal (FF) cell relative to expression of the  
 XX same genes in one or more second filamentous fungal cells. Monitoring  
 XX the global expression of genes from FF cells allows the production  
 XX potential of the microorganisms to be improved. New genes may be  
 XX discovered, possible functions of unknown open reading frames can be  
 XX identified and gene copy number variation and stability can be  
 XX monitored. The expression of genes can be used to study how FF cells

CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organization of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.  
 XX  
 SQ Sequence 1337 BP; 368 A; 318 C; 310 G; 341 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 19.5 Length: 1337  
 Score: 85.00 Matches: 35  
 Percent Similarity: 42.25% Conservative: 25  
 Best Local Similarity: 24.65% Mismatches: 42  
 Query Match: 12.01% Indels: 40  
 DB: 21 Gaps: 7  
 US-10-087-573-2 (1-141) x AAF13411 (1-1337)  
 Qy 2 GluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThr 21  
 Db 273 GAGCAACCGTGATCAAAAGCTACTCAACAGCAGAGCGCCAGCCA----- 317  
 Qy 22 PheAspValMetArgGluAlaLeuLeuArgVallys-----SerSerGluArg 37  
 Db 318 TATGACATCGTACGCGCAGCGCGCTCAAGAGGTTGAACCTGGATAGCTCTCGCGGGTGAA 377  
 Qy 38 -----LeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHis 50  
 Db 378 CTTGAAGACTACGTGGATCTTATCTCCAAAGCTGCGCTACTTCCGGGCAAGCGGG--- 434  
 Qy 51 ArgValLeuProGlyThrGlyAlaSerAlaIleAlaIaThrValThrProLysGlyAla 70  
 Db 435 -----CCTACCGGCACCTGCCAGTCCCGCGGTAGTACCGGTAATGGCGCT 482  
 Qy 71 SerMetLysLeuLysProArgProGlnSerThrLysSerProGluLeuArgGluLeu 90  
 Db 483 GGA-----TCATCTGCCCATGTCTCAAAGGCGAGCATA 515  
 Qy 91 SerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHis 110  
 Db 516 GCGGAGAGAAATTCAC-----GTCCAGGGTTTCATCGGCAAAATGTAATCAT 560  
 Qy 111 ArgLeuProGlyGlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHisLeuArg 130  
 Db 551 ACTATCAATGAG-----GATGAAGGACCGAGTTCACGAGACACATCAAT 605  
 Qy 131 SerLeu 132  
 Db 606 GCCGTT 611  
 RESULT 20  
 AAN60288  
 ID AAN60288 standard; DNA; 9213 BP.  
 AC AAN60288;  
 DT 08-JUN-1991 (first entry)  
 DE Sequence of the HTLV-III genome.  
 KW HIV; LAV; AIDS; diagnosis; vaccine; ss.  
 OS HTLV-III/H9 cells (ATCC CRL 8543).  
 FH Key Location/Qualifiers  
 FT repeat\_region 1..96  
 /\*tag= a

US-10-087-573-2 (1-141) x AAN60288 (1-9213)

QY 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41  
Db 4955 TTGACTGTTTTCAGACTCTCTCTATAAGA----- 4984  
QY 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52  
Db 4985 AAGGCTTTATTAGGACACATAGTTAGCCCTAGGTGTGAATATATCAAGCAGGACATACAAAG 5044  
QY 53 LeuProGlyThrGlyAlaSerAlaAlaThrValThrProLysGlyAlaSerMet 72  
Db 5045 GTAGGATCTCTACATACATTGGCCTAGCAGCATTTAATACACCAAAA----- 5092  
QY 73 LysLeuLysProProArgProGlnSerThrLysSerProGlnLeuArgGluLeuSerArg 92  
Db 5093 AAGATAAGCCACCTTTGCTAGTTTACGAACTGACAGAGGATAGA----- 5140  
QY 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112  
Db 5141 -----TGAACAACAGCCCCAGAGACCAAGGGCCACAGGGGACCAAGATC 5188  
QY 113 ProGluGlyHis--ProLeuLeuGluLys---ArgAlaGluTyrPheArgHisLeuArgS 131  
Db 5189 ---AATGGACACTAGAGCTTTTAGAGGAGCTTAAAGATGAAGCTGTTAGACATTTTCCTA 5245  
QY 131 erLeuLysSerGlnGlyValAsnArgLeuLeu 141  
Db 5246 GGATTGGCTCCATGGCTTAGGGCAACATATC 5277

RESULT 21  
ABV21005/C  
ID ABV21005 standard; cDNA; 1972 BP.  
XX AC ABV21005;  
XX 13-SEP-2002 (first entry)  
DE Human prostate expression marker cDNA 20996.  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX KW Pharmacogenomic marker; gene; ss.  
XX OS Homo sapiens.  
XX PN W0200160860-A2.  
XX 23-AUG-2001.  
XX 20-FEB-2001; 2001WO-US05171.  
XX 17-FEB-2000; 2000US-183319P.  
XX 16-MAR-2000; 2000US-189862P.  
XX 25-MAY-2000; 2000US-207454P.  
XX 09-JUN-2000; 2000US-211314P.  
XX 18-JUL-2000; 2000US-219007P.  
XX 13-DEC-2000; 2000US-255281P.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX PA Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX Claim 1; Page 3457; 11750pp; English.  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:

FT misc\_feature 97...183  
FT /\*tag= b  
FT /label= unique region  
FT CDS 336..731  
FT /\*tag= c  
FT CDS 732..1772  
FT /\*tag= d  
FT /product= p24gag  
FT 1639..4677  
FT /\*tag= e  
FT /product= pol  
FT 4622..5200  
FT /\*tag= f  
FT /product= p'  
FT 5802..7335  
FT /\*tag= g  
FT /product= env  
FT 7336..8373  
FT /\*tag= h  
FT /product= gp41env  
FT 8375..8995  
FT /\*tag= i  
FT /product= E'  
FT 8662..9117  
FT /\*tag= j  
FT /label= unique region  
FT repeat\_region 9118..9213  
FT /\*tag= k  
FT polyA\_signal 9090..9095  
FT /\*tag= l  
FT polyA\_signal 9190..9195  
FT /\*tag= m  
XX EP187041-A.  
XX 09-JUL-1986.  
XX 23-DEC-1985; 85EP-0309454.  
XX 24-DEC-1984; 84US-0685272.  
XX 04-DEC-1985; 85US-0803069.  
XX (GETH ) GENENTECH INC.  
XX Capon DJ, Lasky LA;  
XX P-PSDB; AAP60309, AAP61507, AAP61504, AAP61514, AAP61515.  
XX Acquired immune deficiency syndrome polypeptide(s) - obtd. by  
PT molecular cloning etc. and used for diagnosis and in vaccines  
PT against virus disease  
XX Example; fig 2; 125pp; English.  
XX A comparison of AAN60287 with the cDNA of the HTLV-III genome  
CC revealed one particular clone, designated p7.11 which contained a  
CC DNA sequence encoding this peptide (AAP60308) sequence. This approx.  
CC 2.2 kilobase covers the precursor gag region and encodes, 5' to 3',  
CC p-12, p-15, p-24 a second p-15 protein, and approx. 300 extra base  
CC pairs 3' to the gag region (see AAN60288).  
XX Sequence 9213 BP; 3297 A; 1656 C; 2217 G; 2043 T; 0 other;  
SQ

Alignment Scores:  
Pred. No.: 359 Length: 9213  
Score: 83.50 Matches: 38  
Percent Similarity: 40.91% Conservative: 16  
Best Local Similarity: 28.79% Mismatches: 43  
Query Match: 11.79% Indels: 35  
DB: 7 Gaps: 6

CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 1972 BP; 490 A; 472 C; 557 G; 452 T; 1 other;  
  
Alignment Scores:  
Pred. No.: 53.2 Length: 1972  
Score: 83.00 Matches: 35  
Percent Similarity: 43.01% Conservative: 5  
Best Local Similarity: 37.63% Mismatches: 27  
Query Match: 11.72% Indels: 26  
DB: 23 Gaps: 4  
  
US-10-087-573-2 (1-141) x ABV21005 (1-1972)  
Qy 31 ArgValysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50  
Db 362 CGCACGCTCAGCTCTCCAAAGTTGGTTCCCAACCGCGCAGCTCGGAGTATGCGACATTA 303  
Qy 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaAlaThrVa 65  
Db 302 CCGAGTTGTCTATCATACACGTGGCGCAGCGCGGCGCTCTCTCGCGCAGCCATGTT 243  
Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85  
Db 242 GCGCGCGCTCGAGCC-----CGCAGCGCGCGGCC---GGAACCGCGAGGCC 198  
Qy 85 oGlu-LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluS 105  
Db 197 AGAGCTTCGACGCCCCACAGAGA----- 175  
  
RESULT 22  
ABV26849/C  
ID ABV26849 standard; cDNA; 1972 BP.  
XX AC ABV26849;  
XX DT 16-SEP-2002 (first entry)  
XX DE Human prostate expression marker cDNA 26840.  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX KW pharmacogenomic marker; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200160860-A2.  
XX PD 23-AUG-2001.  
XX PF 20-FEB-2001; 2001WO-US05171.  
XX PR 17-FEB-2000; 2000US-183319P.  
XX PR 16-MAR-2000; 2000US-189862P.  
XX PR 25-MAY-2000; 2000US-207454P.  
XX PR 09-JUN-2000; 2000US-211314P.  
XX PR 18-JUL-2000; 2000US-219007P.  
XX PR 13-DEC-2000; 2000US-255281P.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX Schlegel R, Endege WO, Monahan JE;  
PI WPI; 2001-662795/76.  
DR  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PI for detecting presence of prostate cancer, stage of prostate cancer  
XX  
PS Claim 1; Page 5426; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 1972 BP; 490 A; 472 C; 557 G; 452 T; 1 other;  
  
Alignment Scores:  
Pred. No.: 53.2 Length: 1972  
Score: 83.00 Matches: 35  
Percent Similarity: 43.01% Conservative: 5  
Best Local Similarity: 37.63% Mismatches: 27  
Query Match: 11.72% Indels: 26  
DB: 23 Gaps: 4  
  
US-10-087-573-2 (1-141) x ABV26849 (1-1972)  
Qy 31 ArgValysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50  
Db 362 CGCACGCTCAGCTCTCCAAAGTTGGTTCCCAACCGCGCAGCTCGGAGTATGCGACATTA 303  
Qy 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaAlaThrVa 65  
Db 302 CCGAGTTGTCTATCATACACGTGGCGCAGCGCGGCGCTCTCTCTCGCGCAGCCATGTT 243  
Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85  
Db 242 GCGCGCGCTCGAGCC-----CGCAGCGCGCGGCC---GGAACCGCGAGGCC 198  
Qy 85 oGlu-LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluS 105  
Db 197 AGAGCTTCGACGCCCCACAGAGA----- 175  
  
RESULT 23  
AAS00717  
ID AAS00717 standard; DNA; 1812 BP.  
XX AC AAS00717;  
XX DT 12-SEP-2001 (first entry)  
XX DE Chimeric Caufaq DNA polymerase version 1 polynucleotide.  
XX KW Caufaq DNA polymerase; 3'-5' exonuclease; mutant; proof-reading;  
XX KW second strand DNA synthesis; PCR; ss.

OS Chimeric - Thermus aquaticus.  
XX Chimeric - Unidentified.  
FH Key Location/Qualifiers  
FT 1..1812  
FT /\*tag= a  
FT /product= "Chimeric Cautaq DNA polymerase"  
FT /partial  
FT /note= "No stop codon present"  
XX WO200118213-A1.  
PN 15-MAR-2001.  
PD 08-SEP-2000; 2000WO-GB03478.  
PE 09-SEP-1999; 99GB-0021318.  
PR (DZIE/) DZIEGLEWSKA H.  
PA (KRIS/) KRISTENSEN T.  
XX Kristensen T;  
XX WPI; 2001-218561/22.  
DR P-PSDB; AAU00575.  
XX Chimeric DNA polymerase useful in molecular biology and gene technology  
PT techniques, has domain having proof-reading activity from a DNA  
PT polymerase and DNA polymerase domain from heterologous DNA polymerase -  
PS Claim 13; Page 44-48; 66pp; English.  
XX The sequence represents a polynucleotide encoding a chimeric Cautaq DNA  
CC polymerase polypeptide. The chimeric polypeptide comprises a 3'-5'  
CC exonuclease domain from one DNA polymerase and a DNA polymerase domain  
CC from a heterologous DNA polymerase. This molecule has a domain exhibiting  
CC 3'-5' exonuclease activity derived from Cau DNA polymerase and another  
CC exhibiting DNA polymerase activity from Taq DNA polymerase. These  
CC chimeric sequences are useful in molecular biology and gene technology  
CC techniques, especially second strand DNA synthesis, PCR amplification and  
CC DNA sequencing. The proof-reading (3'-5'exonuclease) activity makes the  
CC polymerase particularly useful in techniques where the fit between the  
CC primer and the template DNA is not exact.  
XX SQ Sequence 1812 BP; 334 A; 574 C; 585 G; 319 T; 0 other;  
Alignment Scores:  
Pred. No.: 53.8 Length: 1812  
Score: 82.50 Matches: 47  
Percent Similarity: 38.01% Conservative: 18  
Best Local Similarity: 27.49% Mismatches: 47  
Query Match: 11.65% Indels: 59  
DB: 22 Gaps: 8  
US-10-087-573-2 (1-141) x AAS00717 (1-1812)  
QY 24 ValMetArgGluAlaLeuLeuArgVallySerSerGluArgLeuAlaMetLeuArgAla 43  
DB 470 ATAAGCAGATTAGCTTGTCTGAGGTGCGGTGAGAGCGGCTCCGTATGCGGCGCG 529  
QY 44 Leu---AlaGlyMetCysGly-HisArgValLeu----- 53  
DB 530 ATGCCCTCTACAGCTTGGGCTCACCAGCGCTTCAGCGCTAGAGCTGAACCTG 589  
QY 53 ----- 53  
DB 590 CGCTGCAGATCTGTATTACAGAGTGAGAGGCCCTTTTCGCTGTCTGCGCCACATGG 649  
QY 54 -----ProGlyThrGlyAlaSerAlaIleAlaIleValThrProLy 68  
DB 650 AGGCCACGGGGGTGGCGCTGACGTGGC-----CTATCTCAGGCGCTTGTCTCT-- 698  
QY 68 sGlyAlaSerMetLysLeuLysProProArg----- 78

699 -GGAGGTGGCGGAGGATCGCCGCTCGAGGCGGAGGTCTTCGCGCTGGCCGCCACC 757  
79 -----ProGln-----SerThrLysSerProGluLeuArgGluLeuSerAr 92  
758 CATTCAACCTCAACTCCGGGACGAGCTGGAAGGGTCTCTTTGACGAGCTAGGGCTTC 817  
92 gLysIleArgGlu-MetAsnLysThrIleSerGlnGlu---SerAlaArgValAsnHisA 111  
818 CGCCCATCGGCAAGACGGAGAGACGGCAAGCGCTCCACGACGCGCGCTCTCTGGAGG 877  
111 rgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHisLeuArgS 131  
878 CCTCCCGGAGGCCACCCCATCTGTGAGAAGATCTTCAGTAC---CGGAGGCTCACCA 934  
131 erLeuLysSerGlnGlyValAsnArgLeu 140  
935 AGCTGAAGACACCTACATTGACCCCTTG 963  
RESULT 24  
AAS00718  
ID AAS00718 standard; DNA; 1812 BP.  
XX AC  
XX AAS00718;  
XX 12-SEP-2001 (first entry)  
DT Chimeric Cautaq DNA polymerase version 2 polynucleotide.  
DE Chimeric Cautaq DNA polymerase version 2 polynucleotide.  
XX Cautaq DNA polymerase; 3'-5' exonuclease; mutant; proof-reading;  
XX second strand DNA synthesis; PCR; ss.  
XX Chimeric - Thermus aquaticus.  
OS Chimeric - Unidentified.  
FH Key Location/Qualifiers  
FT CDS 1..1812  
FT /\*tag= a  
FT /product= "Chimeric Cautaq DNA polymerase"  
FT /partial  
FT /note= "No stop codon present"  
PN WO200118213-A1.  
XX 15-MAR-2001.  
XX 08-SEP-2000; 2000WO-GB03478.  
XX 09-SEP-1999; 99GB-0021318.  
XX (DZIE/) DZIEGLEWSKA H.  
PA (KRIS/) KRISTENSEN T.  
XX Kristensen T;  
XX WPI; 2001-218561/22.  
DR P-PSDB; AAU00574.  
XX Chimeric DNA polymerase useful in molecular biology and gene technology  
PT techniques, has domain having proof-reading activity from a DNA  
PT polymerase and DNA polymerase domain from heterologous DNA polymerase -  
PS Claim 13; Page 52-56; 66pp; English.  
XX The sequence represents a polynucleotide encoding a chimeric Cautaq DNA  
CC polymerase polypeptide. The chimeric polypeptide comprises a 3'-5'  
CC exonuclease domain from one DNA polymerase and a DNA polymerase domain  
CC from a heterologous DNA polymerase. This molecule has a domain exhibiting  
CC 3'-5' exonuclease activity derived from Cau DNA polymerase and another  
CC exhibiting DNA polymerase activity from Taq DNA polymerase. These  
CC chimeric sequences are useful in molecular biology and gene technology  
CC techniques, especially second strand DNA synthesis, PCR amplification and  
CC DNA sequencing. The proof-reading (3'-5'exonuclease) activity makes the  
CC polymerase particularly useful in techniques where the fit between the  
CC primer and the template DNA is not exact.

CC polymerase particularly useful in techniques where the fit between the  
CC primer and the template DNA is not exact.

SQ Sequence 1812 BP; 333 A; 572 C; 587 G; 320 T; 0 other;

Alignment Scores:  
Pred. No.: 53.8 Length: 1812  
Score: 82.50 Matches: 47  
Percent Similarity: 38.01% Conservative: 18  
Best Local Similarity: 27.49% Mismatches: 47  
Query Match: 11.65% Indels: 59  
DB: 22 Gaps: 8

US-10-087-573-2 (1-141) x AA800718 (1-1812)

QY 24 ValMetArgGluAlaLeuArgValLysSerGluArgLeuAlaMetLeuArgAla 43  
DB 470 ATAAGCAGATTAGCTTGTGAGGTGCGGTGACAGCGGCTCGTATCGCGGCGG 529  
QY 44 Leu---AlaGlyMetCysGly-HisArgValLeu----- 53  
DB 530 ATGCCCTCTACAGTTGGGGCTCACCGAGCGCTTGCAGCGTCAGTTAGAGCTGAAGAGA 589  
QY 53 ----- 53  
DB 590 GGCCTCTTGGCTTTATCGGAGGTGGAGAGCGCCCTTTCCGCTGCTCGCCACATGG 649  
QY 54 -----ProGlyThrGlyAlaSerAlaIleAlaIleAlaThrValThrProLy 68  
DB 650 AGGCCACCGGGGTGCGCTCGGACGTGGC-----CTATCTCAGGCGCTTGCCCT-- 698  
QY 68 sGlyAlaSerMetLysLeuLysProProArg----- 78  
DB 699 -GGAGTGGCGGAGGAGATCGCCCGCTCGAGCGCGGTCTTCGCTCGCGGCCAC 757  
QY 79 -----ProGln-----SerThrLysSerProGluLeuArgGluLeuSerAr 92  
DB 758 CTTCAACCTCAACTCCCGGGACCAAGCTGGAAGGGTCTCTTTGACGAGCTAGGGCTTC 817  
QY 92 gLysIleArgGlu-MetAsnLysThrIleSerGlnGlu---SerAlaArgValAsnHisA 111  
DB 818 CCGCCATCGGCAAGACGGAGAGACCGGCAAGCGCTCCACACGCGCCCGCTCTGGAGG 877  
QY 111 rGluProGlyHisProLeuLeuGluLysArgAlaGluTyPheArgHisLeuArgS 131  
DB 878 CCTCCCGAGGCCACCCCATCGTGGGAGAAGATCCTGCAGTAC---CGGAGGCTCACA 934  
QY 131 erLeuLysSerGlnGlyValAsnArgLeu 140  
DB 935 AGCTGAAGAGCACCTACATTGACCCCTTG 963

RESULT 25

AAA82093/C

ID AAA82093 standard; DNA; 480 BP.

AC AAA82093;

DT 04-DEC-2000 (first entry)

XX

DE N. meningitidis partial DNA sequence gnm\_640 SEQ ID NO:640.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

KW antigen; vaccine; diagnosis; infection; antibacterial; identification;

XX Meningococcus B; MenB; ds.

OS Neisseria meningitidis.

XX

XX WO200022430-A2.

PN

PD 20-APR-2000.

XX

PF 08-OCT-1999; 99WO-US23573.

XX

PR 09-OCT-1998; 98US-0103794.  
PR 30-APR-1999; 99US-0132068.

XX (CHIR ) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
Pi Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
Rappuoli R, Pizza M;  
XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
PT used in the diagnosis and treatment of N. meningitidis infection and  
PT other Neisserial infections, for example, N.gonorrhoea -

XX Claim 7; Page 1676; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic  
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
CC represent specifically claimed Neisseria meningitidis genomic DNA  
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
CC sequences, which are all used in the exemplification of the present  
CC invention. The nucleic acid sequences, protein sequences, and antibodies  
CC against them, can be used in the manufacture of a composition. The  
CC composition can be used as a medicament (or in the manufacture of a  
CC medicament) for treating, preventing or diagnosing infection due to  
CC Neisserial bacteria. For example, some of the identified proteins could  
CC be components of vaccines against Meningococcus B; against all serotypes;  
CC and/or against all pathogenic Neisseriae. Identification of sequences  
CC from the bacterium will also facilitate production of biological probes,  
CC particularly organism-specific probes. Attempts to make efficacious  
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
CC Multivalent vaccines have also been tried but none have successfully  
CC overcome antigenic variability. The provision of further, complete  
CC sequences may provide an opportunity to identify secreted or surface  
CC exposed proteins that may be presumed targets for the immune system and  
CC which are not antigenically variable or at least more conserved than  
CC other more variable regions.

XX SQ Sequence 480 BP; 84 A; 141 C; 146 G; 109 T; 0 other;

Alignment Scores:

Pred. No.: 10.6 Length: 480  
Score: 82.00 Matches: 39  
Percent Similarity: 38.85% Conservative: 15  
Best Local Similarity: 28.06% Mismatches: 53  
Query Match: 11.58% Indels: 32  
DB: 21 Gaps: 5

US-10-087-573-2 (1-141) x AAA82093 (1-480)

QY 27 GluAlaLeuArgValLysSerGluArgLeuAlaMetLeuArgAlaLeuAlaGly 46

DB 473 GAGGTGCTGCTGGGCTAAAGAGCGCGGCTAGTAGCGAACCGCGGCTTGAAGCTG 414

QY 47 MetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaIleAlaThrValThr 66

DB 413 ATG-----TGGAGTGTGATGACACACACTGGCGGACCATCTTAGCTCGGTATGATG 360

QY 67 ProLysGlyAlaSerMetLysLeuLysPro-----ProArgPro 79

DB 359 CCCCCAGGTAGGTGCGACGCTTCGAAGCGCGAGAGCGCGGAGGAGCTGCCCTCC 300

QY 80 GlnSerThrLysSerPro----- 85

DB 299 CTGCTGACGACGTGCGCCGAGGGCGCGTGGAGCAATTTGGATCTGAACGCCCGGATG 240

QY 86 GluLeuArgGluLeuSerArgLysIle-----ArgGluMetAsnLysThrIleSer 102

239	GAGCGGAGGCCCTTGAACGCGGAAGATACTTCTGACTAGGAGCACTCCCAAGACCTGGAGG	180	
Db			
Qy	103	GlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArg	122
Db	179	-----GAGCTACGGCTGCTTTGAAGCCGCCAAGATACCCACATTA	138
Qy	123	AlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeuLeu	141
Db	137	ACGCGGCCATTTAGACCCCTTAACACACATCTATCCCTTTACCTCATCAATGAACCTCTTG	81

RESULT 26  
AAS69040  
ID AAS69040 standard; cDNA; 597 BP.

XX  
AC AAS69040;

XX  
DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #4844.

xx Human; chromosome mapping; gene mapping; gene therapy; forensic;  
kw

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX

OS Homo sapiens.  
XX

PN WO200175067-A2.  
XX

PD 11-OCT-2001.  
XX

PF 30-MAR-2001; 2001WO-US08631.  
XX

PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEO INC.

PI Drmarac RT. Liu C. Tang YT;  
XX

XX  
XX  
DR WPI: 2001-639362/73.

DR P-PSDB; ABG04853.  
XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX  
XX  
PS Claim 1, SEQ ID No 4844, 103pp, English

the invention relates to isolated polynucleotide (T) and

the invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers and for chromosome mapping. (II) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers and for chromosome mapping.

CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed  
CC recombinant proteins, and for chromosomal

for identifying expressed genes. (I) is useful in gene therapy techniques for identifying expressed genes. (II) or to treat diseases caused by a deficiency of normal activity of (III) or to treat diseases caused by an overactivity of (IV) are also used in diagnostics as expressed sequence tags. Polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques for identifying expressed genes. (II) or to treat diseases caused by a deficiency of normal activity of (III) or to treat diseases caused by an overactivity of (IV) are also used in diagnostics as expressed sequence tags.

(II). (II) is useful for generating antibodies against it, detecting

quantitating a polypeptide in tissue, as molecular weight markers and  
a food supplement. (II) and its binding partners are useful in medical

CC disorders involving aberrant protein expression or biological activity  
CC imaging of sites expressing (11). (1) and (11) are useful for treating  
CC the polymers and polymers containing polymers have been used in

the polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS6419/-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC  
CC  
CC

CC Note: the sequence data for this patent did not appear in the printee  
CC specification, but was obtained in electronic format directly from WI  
CC

CC at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX

sequence 59/ 59; 9/ A; 21/ C; 196 G; 8/ T; 0 other;

Alignment scores:



PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220563.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0232403.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249267.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-457726/49.

Isolated polypeptide for treating, preventing and prognosing disorders related to the endocrine system including endocrine disorders, reproductive disorders, and gastrointestinal disorders and also for testing and detection e.g. diagnosis -

Disclosure; SEQ ID No 577; 558pp; English.

The invention relates to cDNAs encoding novel human endocrine antigens or a fragment having biological activity, a domain, an epitope, full length protein, variant, allelic variant or a species homologue of the cDNA/antigen. The DNAs and polypeptides are useful for preventing, treating or ameliorating a medical condition when administered (e.g. by gene therapy or antisense-therapy). Identifying mutations in the genes coding for the antigens is useful for diagnosing a pathological condition or a susceptibility to a pathological condition. The DNAs, antigens and antibodies raised against the antigens useful for treating, preventing and/or prognosing disorders related to the endocrine system or hormone imbalance or reproductive disorders, cancers of endocrine

CC tissues, disorders of the pancreas (e.g. diabetes mellitus), the adrenal  
CC glands (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the  
CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples  
CC of diseases and disorders are given in the specification. The present  
CC sequence is genomic DNA fragment form a gene encoding an endocrine  
CC antigen of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 10646 BP; 2501 A; 3544 C; 2986 G; 1615 T; 0 other;

Alignment Scores:  
Pred. No.: 626 Length: 10646  
Score: 82.00 Matches: 23  
Percent Similarity: 45.95% Conservative: 11  
Best Local Similarity: 31.08% Mismatches: 32  
Query Match: 11.58% Indels: 8  
DB: 22 Gaps: 2

US-10-087-573-2 (1-141) x AAS32623 (1-10646)

Qy 46 GlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrVal 65  
Db 5337 GGGGCTACGGGACCGGGTCTGAGTGGGACAGCTGAGGTTTGCACAGGTCACGCTG 5396  
Qy 66 ThrProLysGlyAlaSerMetLysLeuLysProProArgProGln-----Ser 81  
Db 5397 CCACGAGCAGACAGACAGCAGCTGATGTTCCCATGAGGCCCAAGAGCAGAGCTCAGC 5456  
Qy 82 ThrLysSerProGluLeuArgGlu-----LeuSerArgLysIleArgGluMet 97  
Db 5457 ACAGCAGCTCCACTTCGACGAGGGAGAGGAGACAGCCAGGCGGCTTCAAGTTCCTG 5516  
Qy 98 AsnLysThrIleSerGlnGluSerAlaArgValAsnHisArg 111  
Db 5517 CTCTCCAGCTGACGACGACGAGCTCAGAGAGGCCACAGG 5558

RESULT 28  
AAS06738  
ID AAS06738 standard; cDNA; 2838 BP.  
XX AAS06738;  
AC AAS06738;  
DT 12-SEP-2001 (first entry)  
XX Polynucleotide sequence encoding human protein kinase #38.  
DE Human; protein kinase; PTK; STK; cancer; cardiovascular disease;  
KW metabolic disorder; immune related disease; neurological disorder;  
KW neurodegenerative disorder; inflammatory disorder; infectious disease;  
KW reproductive disorder; gene therapy; ss.  
XX Homo sapiens.  
OS  
XX WO200138503-A2.  
PN 31-MAY-2001.  
PD  
XX 22-NOV-2000; 2000WO-US32085.  
PF  
XX 24-NOV-1999; 99US-0167482.  
PR  
XX (SUGEN-) SUGEN INC.  
PA  
XX Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;  
PI Flanagan P, Clary D;  
XX  
DR WPI; 2001-343950/36.  
DR P-PSDB; AAU03538.  
XX

PT Nucleic acids encoding human kinase polypeptides, useful for preventing

PT diagnosing and/or treating e.g. cancer, immune, cardiovascular and  
PT neuronal-associated diseases, and microbial infections -  
XX Example 1; Figure 1; 433pp; English.

XX AAS06701-AAS06757 encode for novel human protein kinases #1-57. The  
CC novel protein kinases have been identified as members of the tyrosine  
CC or serine/threonine kinase (PTK and STK) families. The polynucleotides  
CC encoding protein kinases and the polypeptides may be used in the  
CC prevention, diagnosis and treatment of diseases associated with  
CC cancers (especially cancers of haematopoietic origin), cardiovascular  
CC disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),  
CC immune related diseases (e.g. rheumatoid arthritis), neurological  
CC disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.  
CC Parkinson's disease), inflammatory disorders (e.g. asthma), infectious  
CC disease (e.g. HIV) and reproductive disorders (e.g. infertility).  
CC Additionally, polynucleotides encoding protein kinases may be  
CC used for gene therapy and as DNA probes in diagnostic assays.  
CC The protein kinase polypeptides may be used as antigens in the production  
CC of antibodies against the protein kinases and in assays to identify  
CC modulators of protein kinase expression and activity.

XX Sequence 2838 BP; 694 A; 934 C; 713 G; 497 T; 0 other;

Alignment Scores:  
Pred. No.: 124 Length: 2838  
Score: 81.50 Matches: 30  
Percent Similarity: 41.51% Conservative: 14  
Best Local Similarity: 28.30% Mismatches: 37  
Query Match: 11.51% Indels: 25  
DB: 22 Gaps: 5

US-10-087-573-2 (1-141) x AAS06738 (1-2838)

Qy 43 AlaLeuAlaGlyMetCys-----GlyHisArgValLeuProGly 55  
Db 676 GCCTACCCGGGCTTGTGCGCGCCCGCCACCTGAGTCGGGCGCACCGTTCCTGCCCCCA 735  
Qy 56 ThrGlyAlaSerAla-----IleAlaAlaThr 64  
Db 736 TCGCCCGGCGGCGGCGGTCGCGACCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 795  
Qy 65 ValThrProLysGlyAla-----SerMetLysLeuLysProProArgPro 79  
Db 796 GTGACCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 855  
Qy 80 GlnSerThrLysSerProGluLeuArgGluSerArgLysIleArgGluMetAsnLys 99  
Db 856 CCACCGCCCTCTCCGAAAACATGATACACTTGTATCCCGGATTCACCGGCTGCATCGG 915  
Qy 100 ThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHis---ProLeu 118  
Db 916 AGCAAAATCCACGAG---TTCCAGCTGGGCGCACCGGCTGGAGCGGCGGCGGCGGCG 972  
Qy 119 LeuGluLysArgAlaGlu 124  
Db 973 GCCAAGAAGAAGACAAA 990

RESULT 29  
AAD30562  
ID AAD30562 standard; cDNA; 2962 BP.  
XX AAD30562;  
AC AAD30562;  
DT 21-MAY-2002 (first entry)  
XX Human kinase polypeptide (PKIN-15) cDNA.  
DE Human; kinase polypeptide; PKIN-15; gene therapy; Addison's disease;  
KW leukaemia; immune disorder; lymphoma; melanoma; developmental disorder;  
KW acquired immunodeficiency syndrome; AIDS; allergy; anaemia; hypertension;  
KW asthma; Crohn's disease; rheumatoid arthritis; bursitis; atherosclerosis;

KW cirrhosis; hepatitis; psoriasis; Cushing's syndrome; cytostatic; cancer;  
KW cholestasis; cardiac; cardiovascular disorder; Niemann-Pick's disease;  
KW lipid disorder; fatty liver; Gaucher's disease; myocardial infarction;  
KW drug screening; transgenic animal; antiinflammatory; hepatotropic;  
KW hypotensive; anti-HIV; enzyme; ss.  
XX  
OS Homo sapiens.

XX Key Location/Qualifiers  
FH 85..2921  
FT /\*tag= a  
FT CDS /product= "Human PKIN-15"  
XX  
XX WO200208399-A2.  
XX  
XX 31-JAN-2002.  
XX  
XX 20-JUL-2001; 2001WO-US23092.  
XX  
XX 21-JUL-2000; 2000US-222038P.  
XX  
XX 28-JUL-2000; 2000US-222112P.  
XX  
XX 04-AUG-2000; 2000US-222831P.  
XX  
XX 11-AUG-2000; 2000US-224729P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX (THOR/) THORNTON M.  
XX  
XX Thornton M, Yue H, Khan FA, Gururajan R, Hafalia AJA, Wallia NK;  
XX Patterson C, Ramkumar J, Gandhi AR, Policky JL, Baughn MR;  
XX Triboulet CM, Bandman O, Nguyen DB, Lu Y, Burford N, Lal P;  
XX Ding L, Yao MG, Elliott VS, Recipon SA, Kearney L, Lu DAM;  
XX Greenwald SR, Tang YT, Xu Y, Walsh RT, Gietzen KJ, Yang J;  
XX Hillman JL;  
XX  
XX WPI: 2002-206083/26.  
XX P-PSDB; AA519157.  
XX  
XX New human kinase polypeptide, useful in diagnosis, prevention and  
XX treatment of cancer, immune disorder, growth and developmental  
XX disorder, cardiovascular disorder and lipid disorder  
XX  
XX Claim 5; Page 187-188; 196pp; English.

XX The present invention relates to an isolated human kinase polypeptide  
XX (PKIN) or a biologically active/immunogenic fragment of PKIN. PKIN is  
XX useful for diagnosing, treating and preventing cancer (e.g., leukaemia,  
XX lymphoma, melanoma), an immune disorder (e.g., acquired immunodeficiency  
XX syndrome (AIDS), Addison's disease, allergy, anaemia, asthma, Crohn's  
XX disease, rheumatoid arthritis), a growth and developmental disorder (e.g.  
XX buritis, cirrhosis, hepatitis, psoriasis, Cushing's syndrome), a  
XX cardiovascular disorder (e.g., atherosclerosis, hypertension, myocardial  
XX infarction), and a lipid disorder (e.g., fatty liver, cholestasis,  
XX Gaucher's disease, Niemann-Pick's disease). PKIN is useful in a number of  
XX drug screening techniques and to analyse the proteome of a tissue or cell  
XX type. PKIN is useful for creating knockin humanised animals or transgenic  
XX animals to model human diseases, in somatic or germline gene therapy, to  
XX generate a transcript image of a tissue or cell type, for detecting  
XX differences in the chromosomal location due to translocation, inversion,  
XX etc., among normal, carrier or affected individuals, and as hybridisation,  
XX probes for mapping naturally occurring genomic sequences. PKIN is useful  
XX in southern or northern analysis, dot blot or other membrane-based  
XX technologies, in PCR technologies, in dipstick, pin, microarray enzyme  
XX linked immunosorbent (ELISA)-like assays and in microarrays utilising  
XX fluids or tissues from patients to detect altered PKIN expression. The  
XX present sequence is human PKIN-15 cDNA.

XX Sequence 2962 BP; 713 A; 979 C; 744 G; 526 T; 0 other;

XX Alignment Scores:  
XX Pred. No.: 131 Length: 2962  
XX Score: 81.50 Matches: 30  
XX Percent Similarity: 41.51% Conservative: 14  
XX Best Local Similarity: 28.30% Mismatches: 37

Query Match: 11.51% Indels: 25  
DB: 24 Gaps: 5  
US-10-087-573-2 (1-141) x AAD30562 (1-2962)  
Qy 43 AlaLeuAlaGlyMetCys-----GlyHisArgValLeuProGly 55  
Db 760 GCCTACCCGGCTGTGCGCCGCCGCCCTGAGTGGAGTGGCGGCCCA 819  
Qy 56 ThrGlyAlaSerAla-----IleAlaAlaThr 64  
Db 820 TCGCCCGCGGACGGGACGGGTCCGACCCCGCGGACCCCAACATCGTCACACC 879  
Qy 65 ValThrProLysGlyAla-----SerMetLysLeuLysProProArgPro 79  
Db 880 GTGACCCCGCGGCGACGGCCCATGAGGAAGAAGCTGAAGCCCCCGGGACC 939  
Qy 80 GlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLys 99  
Db 940 CCACCGCCCTCTCCCGAAGAACTGATACACTTCCCGGGATTCCCGCGGTGCATCGG 999  
Qy 100 ThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHis----ProLeu 118  
Db 1000 AGCAATCCCAACGAG---TTCCAGCTGGGGCACCGGTGGACGAGGCCACACGCCCAA 1056  
Qy 119 LeuGluLysArgAlaGlu 124  
Db 1057 GCCAAGAAGAAGACCAA 1074  
RESULT 30  
ABK92234  
ID ABK92234 standard; DNA; 3793 BP.  
XX AC ABK92234;  
XX DT 15-AUG-2002 (first entry)  
XX DE Prostate cancer-associated DNA sequence #120.  
XX KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;  
XX KW gene therapy; gene; ds.  
XX OS Mammalia.  
XX FN WO200230268-A2.  
XX PD 18-APR-2002.  
XX PF 12-OCT-2001; 2001WO-US32045.  
XX PR 13-OCT-2000; 2000US-0687576.  
XX PR 08-DEC-2000; 2000US-0733288.  
XX PR 24-DEC-2000; 2000US-0733742.  
XX PR 24-JAN-2001; 2001US-263957P.  
XX PR 16-MAR-2001; 2001US-276791P.  
XX PR 16-MAR-2001; 2001US-276888P.  
XX PR 06-APR-2001; 2001US-281922P.  
XX PR 24-APR-2001; 2001US-286214P.  
XX PR 30-APR-2001; 2001US-0847046.  
XX PR 04-MAY-2001; 2001US-288589P.  
XX (EOSB-) EOS BIOTECHNOLOGY INC.  
XX PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;  
XX WPI: 2002-471335/50.  
XX DR P-PSDB; ABG61917.  
XX  
XX Detecting a prostate cancer-associated transcript in a cell in a  
XX patient, useful for diagnosing prostate cancer (PC) or screening  
XX modulators of PC, by determining if prostate cancer-associated genes  
XX are expressed in a prostate tissue

XX The present invention relates to methods of detecting a prostate cancer-associated transcript in a cell from a patient. The method comprises contacting a biological sample from the patient with prostate cancer-associated polynucleotides (designated PC genes) that selectively hybridize to a sequence that is at least 80% identical to them. The prostate cancer-associated polynucleotide sequences are differentially expressed in prostate tumour tissue or in prostate cancer and are derived from the tissues of various organisms such as humans or other mammals (e.g. mice, sheep and dogs). The methods of the invention are useful for diagnosing and treating prostate cancer in mammals. The prostate cancer-associated genes are useful for diagnosing or treating prostate cancer, as well as for identifying modulators of prostate cancer or agents that inhibit prostate cancer. The nucleic acid sequences are particularly useful in gene therapy, as a vaccine or in antisense applications. CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide sequences.

XX Sequence 3793 BP; 1269 A; 776 C; 917 G; 831 T; 0 other;

Alignment Scores:  
Pred. No.: 182 Length: 3793  
Score: 81.50 Matches: 37  
Percent Similarity: 39.13% Conservative: 17  
Best Local Similarity: 26.81% Mismatches: 43  
Query Match: 11.51% Indels: 41  
DB: 24 Gaps: 6

US-10-087-573-2 (1-141) x ABK92234 (1-3793)

QY 20 GluThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAla 39  
DB 1963 GAACAATTGGATGCCATCAACAAGAAATCAGGCTAATTCAGGAAGAAAAGAAATCTACA 2022  
QY 40 MetLeuArgAla-----LeuAlaGlyMetCys 48  
DB 2023 GAGTTCGGTGCTGAGAAATTCAGAAATAGAGTGGCTAGTGTGAGCTCGAAGGCTGAAT 2082  
QY 49 GlyHisArgValLeuProGlyThrGly-----AlaSerAlaIleAla 62  
DB 2083 TTGGCAAGGGTCACCCAGGTACCTCCATTACTGCTCTGTTACAGCTTCATCGTGCC 2142  
QY 63 AlaThrValThrProLysGlyAlaSerMet---LysLeuLysProProArgProGln--- 80  
DB 2143 AGTTTCATCTCCCGCCAGTCAGTGGACACTCAACTCCAAAGCTCACCCCTCGAAGCCCTGCCAGG 2202  
QY 81 -----SerThrLysSerProGluLeuArgGluLeuSerArgLys 93  
DB 2203 GAAATGGATCGGATCGGAGTCATGACACTGCCAAGTGTATCTGAGAAACATCGGAGAAAG 2262  
QY 94 Ile-----ArgGluMetAsnLysThrIleSerGlnGluSerAla 106  
DB 2263 ATTGCAGTTGTGGAAGAAGATGTCGAGAGGACAAAGCAACAATTAATGTGAACACTTCT 2322  
QY 107 -----ArgValAsnHisArgLeuProGluGlyHis 116  
DB 2323 CCTCTCTCTACCCCTAGAGCCCTCAGATGACTCAGCTCTCCCTTCTTCCTAC 2376

RESULT 31  
ID AAT14180  
XX AAT14180 standard; cDNA; 9718 BP.

AC AAT14180;

XX 03-JUL-1996 (first entry)

XX Attenuated HIV-1 strain CpG1 genome.

XX HIV-1; human immunodeficiency virus type 1; AIDS; attenuation;  
KW vaccine; DNA methylation; ss.

XX

OS Synthetic.

XX WO9611280-A1.  
XX 18-APR-1996.  
XX 05-OCT-1995; 95WO-US13219.  
XX 07-OCT-1994; 94US-0319974.  
XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI; 1996-209861/21.

XX Attenuated virus contg. additional methylation sites in its genome  
PT - used to induce protective immunity and to raise antibodies for  
PT diagnostic use.

XX Example 1; Page 23-28; 40pp; English.

XX A novel, attenuated HIV-1 genome, HIV-1 CpG1 (AAT14180), is  
CC hyperubstituted with noninformational or 'silent' CpG segments.  
CC Addition of the CpGs to the genome does not alter the amino acid  
CC sequence of the encoded proteins but makes the synthetic genome a  
CC target for host cell-methylases. Thus, although the virus for which  
CC this genome codes is capable of infecting a cell, the proviral genome  
CC is easily inactivated by methylation and kept permanently in a dormant  
CC state. The virus can be used for the prepn. of live virus vaccine or  
CC to raise diagnostic antibodies.

XX Sequence 9718 BP; 2797 A; 2198 C; 2771 G; 1952 T; 0 other;

Alignment Scores:  
Pred. No.: 628 Length: 9718  
Score: 81.50 Matches: 34  
Percent Similarity: 37.50% Conservative: 11  
Best Local Similarity: 28.33% Mismatches: 38  
Query Match: 11.51% Indels: 37  
DB: 17 Gaps: 6

US-10-087-573-2 (1-141) x AAT14180 (1-9718)

QY 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41  
DB 5373 TTGACTGTTTTCGGACTCGGCGATACGC-----5402  
QY 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52  
DB 5403 AAGCGGTACTCGGACACATCGTTTCGCCGCTGCGAATATCAAGCGGACATACAAG 5462  
QY 53 LeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMet 72  
DB 5463 GTCGATCGCTACATAACCTCGCGCTCGCGCGCTTAATACGCGGAAA-----5510  
QY 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92  
DB 5511 AAGATAAAGCGCGCGTTTCGCGTTCGAAACTGACGAGGATCGA-----5558  
QY 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112  
DB 5559 -----TGGAAACAAGCCCGGAGAGACAGAGCCGACAGAGGAGGACCAATG 5606  
QY 113 ProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArg 127  
DB 5607 ---AATGACACTAGAGCTTTTAGAGAGGCTTAAAGAACGAGCGGTTCGCCATTTTCGCCG 5663

RESULT 32

ABL28661/c

ID ABL28661 standard; DNA; 1126 BP.

XX ABL28661;

XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37456.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical; gene; ds.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US09231.  
XX PR 23-MAR-2000; 2000US-191637P.  
XX PR 11-JUL-2000; 2000US-0614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX PI WPI; 2001-656860/75.  
XX DR New isolated nucleic acid detection reagent for detecting 1000 or more  
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX PT interactions -  
XX PS Claim 1; SEQ ID NO 37456; 21pp + Sequence Listing; English.  
XX CC The invention relates to an isolated nucleic acid detection reagent  
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
XX CC useful in developmental biology and in elucidating cell signalling and  
XX CC cell-cell interactions in higher eukaryotes for the development of  
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention  
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX CC sequences (ABL01840-ABL16175) and the encoded proteins  
XX CC (ABB57737-ABB72072).  
XX CC The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 1126 BP; 263 A; 320 C; 340 G; 203 T; 0 other;

Alignment Scores:  
Pred. No.: 41.6 Length: 1126  
Score: 81.00 Matches: 43  
Percent Similarity: 43.48% Conservative: 27  
Best Local Similarity: 26.71% Mismatches: 58  
Query Match: 11.44% Indels: 34  
DB: 23 Gaps: 7

US-10-087-573-2 (1-141) x ABL28661 (1-1126)  
QY 8 ThrAsnPhaValAlaGluAsnArgProThrPheGly----- 19  
Db 714 TCGACATGTTGCCCTCGACCTCTGTTGTTGGTGTGCAACGTGCAGAAATCCCAT 655  
QY 20 -----GluThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArg 37  
Db 654 TGTCTTCAGCATCTCGACCTCTCGCGCATTTACCTCCAGGTC-----CGC 607  
QY 38 LeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuPro----- 54  
Db 606 CTCGTTGATCTTGGCCCGGTC---GCGAGCTGTGCGCGCGCATTTCTTCGTGCTCCT 550  
QY 55 -----GlyThrGlyAlaSerAlaAlaAlaThrValThrProLysGlyAla 70  
Db 549 TCGCGACGTCGATGTTCCGAGCTGCTGTTGCGCGCGTCTTCGCGCGCGCTCCTCC 490  
QY 71 SerMetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeu 90

Db 489 TCCTCTCTCTCCGGGACCCTCGACCGCATCT---AGATCGCGA--CTCGGCGCACGG 435  
QY 91 SerArg-----LysIleArgGluMetAsnLysThrIle 101  
Db 434 TCGCGCTCCCGTTCCCGCTCCCGATCTCGGCTGGGACTCGCGTCCGATCCGCTTCCCGG 375  
QY 102 SerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLys 121  
Db 374 TCGCGGACCGCGCGCTCCAGATCCGATCTCGGCTGGGCTACTCCGATGCGCTCC 315  
QY 122 ArgAlaGluTyrPheArgHisLeuArgSerLeuLysSer---GlnGlyValAsnArgLeu 140  
Db 314 CGTCCCGGGAACGGCGCTCCCTCGCTTCTTCGCTCGGATCGCTCTCTTCGCGTCTC 255  
QY 141 Ile 141  
Db 254 CTG 252  
RESULT 33  
ID AAH77199 standard; DNA; 1797 BP.  
XX AAH77199;  
XX AC AAH77199;  
XX DT 24-JAN-2002 (first entry)  
XX DE Phaeotrichoconis crotalariae haloperoxidase gene.  
XX KW Haloperoxidase; enzyme; halide ion; bactericide; detergent; disinfectant;  
XX KW microbial growth inhibition; preservation agent; enzymatic bleaching; ds.  
XX OS Phaeotrichoconis crotalariae.  
XX FH Location/Qualifiers  
XX CDS 1..1797  
XX FT /\*tag= a  
XX FT /product= "Haloperoxidase"  
XX FT /partial  
XX FT /note= "No stop codon given"  
XX PN WO200179462-A2.  
XX PD 25-OCT-2001.  
XX PF 10-APR-2001; 2001WO-DK00245.  
XX PR 14-APR-2000; 2000DK-0000625.  
XX PA (NOVO ) NOVOZYMES AS.  
XX PA (MAXY-) MAXYGEN INC.  
XX PI Danielssen S, Schneider P;  
XX DR WPI; 2002-026023/03.  
XX DR P-PSDB; AAG77921.  
XX PT Isolated nucleic acids which encode polypeptides having haloperoxidase  
XX PT activity which can be used in compositions for anti-bacterial activity  
XX BS Claim 1; Page 47-49; 51pp; English.  
XX CC The sequence represents the novel polynucleotide of the invention  
XX CC encoding a polypeptide having haloperoxidase activity, isolated from  
XX CC Phaeotrichoconis crotalariae. The haloperoxidase of the invention works  
XX CC as an enzyme or by oxidation of halide ions, having bactericidal  
XX CC activity. The nucleotide sequence is useful for DNA shuffling  
XX CC (recombination) such that new polynucleotide sequences obtained may  
XX CC encode new polypeptides having haloperoxidase activity with improved  
XX CC properties. The invention relates to methods for killing or inhibiting  
XX CC the growth of microbial cells. The haloperoxidase of the invention may be  
XX CC added to become a component of a detergent composition. The compositions  
XX CC may also comprise one or more other enzymes such as a protease, a lipase,

CC a cutinase or an amylase. The haloperoxidases may be used as preservation  
 CC agents and disinfectant agents. The haloperoxidase also may be used in  
 CC enzymatic bleaching applications.  
 XX

SQ Sequence 1797 BP; 379 A; 584 C; 473 G; 361 T; 0 other;

Alignment Scores:
 Pred. No.: 77 Length: 1797
 Score: 81.00 Matches: 27
 Percent Similarity: 38.64% Conservative: 7
 Best Local Similarity: 30.68% Mismatches: 27
 Query Match: 11.44% Indels: 27
 DB: 24 Gaps: 3
 US-10-087-573-2 (1-141) x AAH77199 (1-1797)

QY 54 ProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLys 73  
 DB 1073 CCGGCTACGGATGACGGCGCTCAGACACCGGCGACCCCTTCTGCTCACCTCGCG 1132  
 QY 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89  
 DB 1133 CTCTGCGCAACACAAACAGACATCCCTTCAAGCCCCCTTCCCGCTACCGGCG 1192  
 QY 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109  
 DB 1193 GCCACGGACCTTTGGCGG----- 1211  
 QY 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121  
 DB 1212 -CGCGCTCTTCCAATGGTGGCGGCTACTACACGGCGGCGTGGCAGCGTGGAGGACA 1270  
 QY 122 -ArgAlaGluTyrPheArgHis 128  
 DB 1271 ACGAGCGCGACAAACATCGCCAT 1292

RESULT 34  
 AAH77155  
 ID AAH77155 standard; DNA; 1797 BP.  
 XX AC AAH77155;  
 XX 21-JAN-2002 (first entry)  
 XX Haloperoxidase coding sequence.

XX Haloperoxidase; antibacterial; microbial growth inhibition; preservation;  
 KW disinfectant; enzymatic bleaching; pulp bleaching; stain bleaching; ds.  
 OS Phaeotrichoconis crotalariae.  
 XX Key Location/Qualifiers  
 FT CDS 1..1797  
 FT /\*tag= a  
 FT /product= "Haloperoxidase"  
 FT /partial  
 FT /note= "No stop codon given"

XX W0200179461-A2.  
 XX 25-OCT-2001.  
 XX 10-APR-2001; 2001WO-DK00243.  
 XX 14-APR-2000; 2000DK-0000625.  
 XX (NOVO ) NOVOZYMES A/S.  
 XX Danielsen S, Schneider P;  
 XX WPI; 2002-017611/02.  
 XX P-ESDB; AAG77905.  
 XX

PT New haloperoxidase purified from Phaeotrichoconis crotalariae has  
 PT bactericidal activity and is useful as a preservative or disinfectant,  
 PT for example in body care products and food preparation -  
 XX

PS Claim 1; Page 46-48; 50pp; English.

CC The sequence represents the coding sequence for the novel polypeptide  
 CC of the invention having haloperoxidase activity. The polypeptide of the  
 CC invention also has antibacterial activity. The haloperoxidase is used to  
 CC kill or inhibit growth of microbial cells. It may be used to preserve or  
 CC disinfect, for example in water based paints or personal care products,  
 CC for cleaning surfaces and cooking utensils in food processing plants and  
 CC any area where food is prepared or served. It may also be used in  
 CC enzymatic bleaching applications, for example pulp or stain bleaching.  
 XX

SQ Sequence 1797 BP; 379 A; 584 C; 473 G; 361 T; 0 other;

Alignment Scores:
 Pred. No.: 77 Length: 1797
 Score: 81.00 Matches: 27
 Percent Similarity: 38.64% Conservative: 7
 Best Local Similarity: 30.68% Mismatches: 27
 Query Match: 11.44% Indels: 27
 DB: 24 Gaps: 3
 US-10-087-573-2 (1-141) x AAH77155 (1-1797)

QY 54 ProGlyThrGlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLys 73  
 DB 1073 CCGGCTACGGATGACGGCGCTCAGACACCGGCGACCCCTTCTGCTCACCTCGCG 1132  
 QY 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89  
 DB 1133 CTCTGCGCAACACAAACAGACATCCCTTCAAGCCCCCTTCCCGCTACCGGCG 1192  
 QY 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109  
 DB 1193 GCCACGGACCTTTGGCGG----- 1211  
 QY 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121  
 DB 1212 -CGCGCTCTTCCAATGGTGGCGGCTACTACACGGCGGCGTGGCAGCGTGGAGGACA 1270  
 QY 122 -ArgAlaGluTyrPheArgHis 128  
 DB 1271 ACGAGCGCGACAAACATCGCCAT 1292

RESULT 35  
 ABK84630/C  
 ID ABK84630 standard; cDNA; 3530 BP.  
 XX AC ABK84630;  
 XX 14-AUG-2002 (first entry)  
 XX Human cDNA differentially expressed in granulocytic cells #1201.

KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; periodontal disease;  
 KW granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.  
 XX WO200228999-A2.  
 XX 11-APR-2002.  
 XX 03-OCT-2001; 2001WO-US30821.

03-OCT-2000; 2000US-237189P.  
(GENE-) GENE LOGIC INC.  
Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
WPI; 2002-435328/46.  
Detecting granulocyte activation by detecting differential expression of genes associated with granulocyte activation, which serves as diagnostic markers that is useful for monitoring disease states and drug toxicity -  
Claim 1; SEQ ID No 1201; 114pp; English.  
The invention relates to detecting (M1) granulocyte (GC) activation (GCA), by detecting the level of expression of gene(s) (Gs) identified by DNA chip analysis as given in the specification, and comparing the expression level to an expression level in an unactivated GC, where differential expression of Gs is indicative of GCA. Also included are modulating (M2) GA by contacting GC with an agent that alters the expression of at least one gene in Gs; (2) screening (M3) for an agent capable of modulating GCA or an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection, parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from Wipo at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
Sequence 3530 BP; 735 A; 902 C; 1044 G; 848 T; 0 other;

```

Alignment Scores:
Pred. No.:      187      Length:      3530
Score:          81.00    Matches:      30
Percent Similarity: 38.04% Conservative: 5
Best Local Similarity: 32.61% Mismatches: 30
Query Match:      11.44% Indels:      25
DB:               24      Gaps:       2

US-10-087-573-2 (1-141) x ABX84630 (1-3530)

Qy      31  ArgVallySerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50
      |||||
Db      595  CGCACGCTCAGTCTCCAGGTTCCTCCACCGCGGCATCGGGAGTATGACAGATT 536

Qy      50  sArgValLeu-----ProGlyThrGlyAlaSerAlaIleAlaThrVa 65
      :|||:::
Db      535  CCAGTGTGTTCATCATACATACAGTCGCGGCCAGGCGGGGGCTCTCTCTCGCGGCAGCCATGTT 476

```

Qy	65	lthrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr	85
Db	475	GGCGCCCGTCGGAGCCGCGACGCCGCCCGGAACCCGAGCCCGAGCCCTTCGACGCC	416
Qy	85	oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe	105
Db	415	CCAG-----	412
Qy	105	rAlaArgValAsnHisArgLeuProGluGlyHis	116
Db	411	-AGACGGGCTGGCGCGCGCGAACCAGAGGCCAC	379
RESULT 36			
ABN97276/c			
ID	ABN97276	standard; DNA; 3530 BP.	
XX	ABN97276;		
XX			
DT	13-AUG-2002	(first entry)	
XX			
DE	Gene #3774	used to diagnose liver cancer.	
XX			
KW	Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;		
KW	metastatic liver tumour; cytostatic; expression profile; disease state;		
KW	disease progression; drug toxicity; drug efficacy; drug metabolism.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200229103-A2.		
XX			
PD	11-APR-2002.		
XX			
PF	02-OCT-2001; 2001WO-US30589.		
XX			
PR	02-OCT-2000; 2000US-237054P.		
XX			
FA	(GENE-) GENE LOGIC INC.		
XX			
FI	Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;		
XX			
DR	WPI; 2002-426119/45.		
XX			
PT	Diagnosing and detecting the progression of liver cancer.		
PT	hepatocellular carcinoma or metastatic liver tumor in a patient,		
PT	involves detecting the level of expression of two or more genes in a		
PT	liver tissue sample -		
XX			
PS	Claim 1; SEQ ID NO 3774; 298pp; English.		

CC The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumour in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in A8993503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytostatic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published/pct\\_sequences](http://ftp.wipo.int/pub/published/pct_sequences).

XX	Sequence	3530 BP; 736 A; 902 C; 1044 G; 848 T; 0 other;	
SQ			
	Alignment Scores:		
	Pred No.:	187	Length: 3530
	Score:	81.00	Matches: 30
	Percent Similarity:	38.04%	Conservative: 5
	Best Local Similarity:	32.61%	Mismatches: 25
	Query Match:	11.44%	Indels: 32



DB: 24 Gaps: 2  
US-10-087-573-2 (1-141) x ABL69227 (1-3530)  
Qy 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50  
Db 595 CGCACGCTCAGCTCTCCAAAGGTTGGCTTCCCCACCGCGCACTCGGAGATGACGACATTA 536  
Qy 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaThrVa 65  
Db 535 CCGAGTTGTCTATCATACAGTCGCGCAGCGCGGGCTCTCTCGCGCAGCCATGTT 476  
Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85  
Db 475 GCGCGCGCTCGGAGCGCGCA3CCCGCGCGCGGAAACCGGAGCCCGACAGCCCTTCGACGCC 416  
Qy 85 oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe 105  
Db 415 CCAG----- 412  
Qy 105 rAlaArgValAsnHisArgLeuProGluGlyHis 116  
Db 411 -AGACGGCTGGCGCGCGCGGAAACCGAGGGCCAC 379  
RESULT 37  
ABL69227/C  
ID ABL69227 standard; DNA; 3530 BP.  
XX AC ABL69227;  
XX DT 15-MAY-2002 (first entry)  
XX DE Prostate cancer related gene sequence SEQ ID NO: 7564.  
XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytosol; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX OS Homo sapiens.  
XX PN WO200194629-A2.  
XX PD 13-DEC-2001.  
XX PF 30-MAY-2001; 2001WO-US10838.  
XX PR 05-JUN-2000; 2000US-208473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 22-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.

28-SEP-2000; 2000US-236111P.  
29-SEP-2000; 2000US-236842P.  
29-SEP-2000; 2000US-236891P.  
02-OCT-2000; 2000US-237172P.  
02-OCT-2000; 2000US-237173P.  
02-OCT-2000; 2000US-237278P.  
02-OCT-2000; 2000US-237294P.  
02-OCT-2000; 2000US-237295P.  
02-OCT-2000; 2000US-237316P.  
03-OCT-2000; 2000US-237425P.  
03-OCT-2000; 2000US-237598P.  
03-OCT-2000; 2000US-237604P.  
03-OCT-2000; 2000US-237606P.  
03-OCT-2000; 2000US-237608P.  
01-NOV-2000; 2000US-244867P.  
01-NOV-2000; 2000US-245084P.  
XX (AVAL-) AVALON PHARM.  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Sopet DR, Weaver Z;  
XX WPI; 2002-188264/24.  
XX  
XX Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set  
PS Claim 1; SEQ ID 7564; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 847 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
SQ Sequence 3530 BP; 736 A; 902 C; 1044 G; 848 T; 0 other;  
Alignment Scores:  
Pred. No.: 187 Length: 3530  
Score: 81.00 Matches: 30  
Percent Similarity: 38.04% Conservative: 5  
Best Local Similarity: 32.61% Mismatches: 32  
Query Match: 11.44% Indels: 25  
DB: 24 Gaps: 2  
US-10-087-573-2 (1-141) x ABL69227 (1-3530)  
Qy 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50  
Db 595 CGCACGCTCAGCTCTCCAAAGTGGCTTCCCCACCGCGCACTCGGAGATGACGACATTA 536  
Qy 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaThrVa 65  
Db 535 CCGAGTTGTCTATCATACAGTCGCGCAGCGCGGGCTCTCTCGCGCAGCCATGTT 476  
Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85  
Db 475 GCGCGCGCTCGGAGCGCGCGGAAACCGGAGCCCGACAGCCCTTCGACGCC 416  
Qy 85 oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe 105

Db 415 CCAG----- 412

Qy 105 rAlaArgValaenHisArgLeuProGluGlyHis 116  
|||||

Db 411 -AGACGGGCTGGCGCGCCGACACCGAGGGCCAC 379  
|||||

RESULT 38  
ABL69756/c  
ID ABL69756 standard; DNA; 3530 BP.

XX ABL69756;  
AC  
XX 15-MAY-2002 (first entry)  
XX  
XX Prostate cancer related gene sequence SEQ ID NO:8093.  
XX  
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200194629-A2.  
PN  
XX 13-DEC-2001.  
PD  
XX 30-MAY-2001; 2001WO-US10838.  
XX  
XX 05-JUN-2000; 2000US-209473P.  
PR  
XX 05-JUN-2000; 2000US-209531P.  
PR  
XX 18-SEP-2000; 2000US-233133P.  
PR  
XX 18-SEP-2000; 2000US-233617P.  
PR  
XX 20-SEP-2000; 2000US-234009P.  
PR  
XX 20-SEP-2000; 2000US-234034P.  
PR  
XX 20-SEP-2000; 2000US-234052P.  
PR  
XX 22-SEP-2000; 2000US-234099P.  
PR  
XX 22-SEP-2000; 2000US-234567P.  
PR  
XX 25-SEP-2000; 2000US-234923P.  
PR  
XX 25-SEP-2000; 2000US-234924P.  
PR  
XX 25-SEP-2000; 2000US-235077P.  
PR  
XX 25-SEP-2000; 2000US-235082P.  
PR  
XX 25-SEP-2000; 2000US-235134P.  
PR  
XX 25-SEP-2000; 2000US-235280P.  
PR  
XX 26-SEP-2000; 2000US-235637P.  
PR  
XX 26-SEP-2000; 2000US-235638P.  
PR  
XX 27-SEP-2000; 2000US-235711P.  
PR  
XX 27-SEP-2000; 2000US-235720P.  
PR  
XX 27-SEP-2000; 2000US-235840P.  
PR  
XX 27-SEP-2000; 2000US-235863P.  
PR  
XX 28-SEP-2000; 2000US-236028P.  
PR  
XX 28-SEP-2000; 2000US-236032P.  
PR  
XX 28-SEP-2000; 2000US-236033P.  
PR  
XX 28-SEP-2000; 2000US-236034P.  
PR  
XX 28-SEP-2000; 2000US-236109P.  
PR  
XX 28-SEP-2000; 2000US-236111P.  
PR  
XX 29-SEP-2000; 2000US-236842P.  
PR  
XX 29-SEP-2000; 2000US-236891P.  
PR  
XX 02-OCT-2000; 2000US-237172P.  
PR  
XX 02-OCT-2000; 2000US-237173P.  
PR  
XX 02-OCT-2000; 2000US-237278P.  
PR  
XX 02-OCT-2000; 2000US-237294P.  
PR  
XX 02-OCT-2000; 2000US-237295P.  
PR  
XX 02-OCT-2000; 2000US-237316P.  
PR  
XX 03-OCT-2000; 2000US-237425P.  
PR  
XX 03-OCT-2000; 2000US-237598P.  
PR  
XX 03-OCT-2000; 2000US-237604P.  
PR  
XX 03-OCT-2000; 2000US-237606P.  
PR  
XX 03-OCT-2000; 2000US-237608P.  
PR  
XX 01-NOV-2000; 2000US-244867P.  
PR  
XX 01-NOV-2000; 2000US-245084P.  
XX

PA (AVAL-) AVALON PHARM.  
XX  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX MPI; 2002-188264/24.  
XX  
XX Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set  
XX  
XX Claim 1; SEQ ID 8093; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 847 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
XX Sequence 3530 BP; 736 A; 902 C; 1044 G; 848 T; 0 other;  
SQ

Alignment Scores:  
Pred. No.: 187 Length: 3530  
Score: 81.00 Matches: 30  
Percent Similarity: 38.04% Conservative: 5  
Best Local Similarity: 32.61% Mismatches: 32  
Query Match: 11.44% Indels: 25  
DB: 24 Gaps: 2

US-10-087-573-2 (1-141) x ABL69756 (1-3530)

Qy 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50  
|||

Db 595 CGACGGCTCAGCTCTCCAGGTGGCTTCCCGCCGCGGCTCTCTCGCGGCGCATGTT 536  
|||

Qy 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaAlaThrVa 65  
|||

Db 535 CCGAGTTGTCATCATACACGTCGCGCGGCGGCGGCTCTCTCGCGGCGCATGTT 476  
|||

Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85  
|||

Db 475 GGCGCCCGTCGAGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 416  
|||

Qy 85 oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe 105  
|||

Db 415 CCAG----- 412

Qy 105 rAlaArgValAsnHisArgLeuProGluGlyHis 116  
|||

Db 411 -AGACGGGCTGGCGCGCCGAAACCGAGGGCCAC 379  
|||

RESULT 39  
ABL26682/c  
ID ABL26682 standard; DNA; 6105 BP.  
XX  
XX ABL26682;  
AC  
XX 26-MAR-2002 (first entry)  
DT  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 31519.  
DE  
XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
OS Drosophila melanogaster.  
PN WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-655860/75.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX Claim 1; SEQ ID NO 31519; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (BB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 6105 BP; 1624 A; 1403 C; 1377 G; 1701 T; 0 other;

Alignment Scores:  
Pred. No.: 385 Length: 6105  
Score: 81.00 Matches: 33  
Percent Similarity: 39.32% Conservativity: 13  
Best Local Similarity: 28.21% Mismatches: 31  
Query Match: 11.44% Indels: 40  
DB: 23 Gaps: 4

US-10-087-573-2 (1-141) x ABL26682 (1-6105)  
Qy 20 GluThrPheaspValMetArg-----GluAlaLeuLeuArg 31  
Db 4274 GAAGATTTCAGCCCATGACGAGCAGGAGATCTGAAATGCGCATCGAGAGATGGTTAGA 4215  
Qy 32 Vallys---SerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHis 50  
Db 4214 ATCAGGTGTAGTACCGAGAGACTCTATAATAGT----- 4179  
Qy 51 ArgValLeuProGlyThrGlyAlaSerAlaAlaThrValThrProLysGlyAla 70  
Db 4179 ----- 4179  
Qy 71 SerMetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeu 90  
Db 4178 AGCCAGAGAGCTCAACCTCTCTCATCAGACACCGTGTCCGTGGAGTTCGCGGATTTG 4119  
Qy 91 SerArgLysLeuArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHis 110  
Db 4118 GAGTCCGCTGCCAGAGAGCTGCAATTCGACATTAGCCAGTACACCAAGAGATCGAGGAG 4059  
Qy 111 ArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyrPheArg 127  
Db 4058 CTCACACAGGAA-----CTGCTTAGGAGGCAACACAGAACTGGAGCCG 4017

RESULT 40  
ABQ51890/c  
ID ABQ51890 standard; DNA; 681 BP.  
XX  
AC ABQ51890;  
XX 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 38481.  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX Homo sapiens.  
XX WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.  
XX (EPIC-) EPIGENOMICS AG.  
PA Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ34410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX SQ Sequence 681 BP; 67 A; 83 C; 251 G; 280 T; 0 other;  
Alignment Scores:  
Pred. No.: 24.3 Length: 681  
Score: 80.50 Matches: 23  
Percent Similarity: 41.03% Conservativity: 9  
Best Local Similarity: 25.49% Mismatches: 35  
Query Match: 11.37% Indels: 11  
DB: 24 Gaps: 2  
US-10-087-573-2 (1-141) x ABQ51890 (1-681)  
Qy 67 ProLysGlyAlaSerMetLysLeuLysPro-----ProArgPro 79

```

Db      ||||| ||| ::||| ||||| |||||
450 CCTAAACGCGATATAAAACGAAACCGACCGCTACCCCTCCAAACGCCTCGACCG 391
Qy      80 GlnSerThrLysSerProGluLeu-----ArgGluLeuSerArgLysIleArg 95
Db      ::||| ||||| ||||| ||||| ||||| |||||
390 CCAATACCCCGAAACCGACCTCCCTCCACCCCGCGCTATCCCGACGCAACCG 331
Qy      96 GluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGly 115
Db      ||||| ||||| ::||| ::|||
330 AACCGTAACAATCAGCTCTCCCATCCGACCCCAATAAAACCAACCGCAACCGCGCC 271
Qy      116 HisProLeuLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLys 133
Db      ||||| ||||| ::||| ||||| |||||
270 GACCCCGCGCTAATATAAAACGCGGAAAAAACGAAAAAACGAAAAACAAAAA 217

```

Search completed: November 17, 2003, 13:04:56  
 Job time : 280 secs

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 17, 2003, 12:56:54 : Search time 56 Seconds

(without alignments)  
1111.340 Million cell updates/sec

Title: US-10-087-573-2

Perfect score: 708

Sequence: 1 MESTSTTTTFVAENRPTFGE.....RAEYFRHLRLSLKSGVNRLLI 141

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US10087573/tunat 14112003 103611 20707/app query.fasta\_1.327  
-DB=Issued Patents NA -QFWT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOPECL=0  
-LCOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10087573@cgn 1 1 44 @tunat 14112003 103611 20707 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	13.0	774	4	US-09-252-991A-15297
2	92	13.0	2163	4	US-09-252-991A-15344
3	92	13.0	4260	4	US-09-252-991A-15251
C 4	84.5	11.9	1530	4	US-09-252-991A-6770
5	84.5	11.9	1695	4	US-09-252-991A-6969
C 6	83	11.7	1677	4	US-09-252-991A-4734
7	83	11.7	2151	4	US-09-252-991A-4774
C 8	82.5	11.7	750	4	US-09-252-991A-12401
9	82.5	11.7	1581	4	US-09-252-991A-12190
10	82.5	11.7	1620	4	US-09-252-991A-12255
11	81	11.4	1797	4	US-09-833-441-1
12	81	11.4	1797	4	US-09-833-102-1

13	80	11.3	1422	4	US-09-252-991A-14835	Sequence 14835, A
14	80	11.3	1446	4	US-09-252-991A-15091	Sequence 15091, A
C 15	79.5	11.2	1149	3	US-08-660-645A-8	Sequence 8, Appli
C 16	79.5	11.2	1149	3	US-09-298-718-8	Sequence 8, Appli
C 17	79.5	11.2	1149	3	US-09-546-969-8	Sequence 8, Appli
18	79.5	11.2	8625	3	US-08-980-832-1	Sequence 1, Appli
19	79.5	11.2	11233	3	US-08-980-832-27	Sequence 27, Appli
C 20	78.5	11.1	997	4	US-09-452-239-13	Sequence 13, Appli
21	78.5	11.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
22	78.5	11.1	4415229	3	US-09-103-840A-1	Sequence 1, Appli
23	78	11.0	2061	4	US-09-107-532A-1348	Sequence 1348, Ap
24	77.5	10.9	3856	2	US-07-743-357-20	Sequence 20, Appli
25	77.5	10.9	4132	2	US-07-743-357-21	Sequence 21, Appli
26	77.5	10.9	8932	3	US-09-124-900-1	Sequence 1, Appli
27	77.5	10.9	8933	3	US-08-463-210-4	Sequence 4, Appli
28	77.5	10.9	8933	3	US-09-620-958A-3	Sequence 3, Appli
29	77.5	10.9	8933	3	US-09-620-958A-4	Sequence 4, Appli
30	77.5	10.9	8933	3	US-09-620-958A-9	Sequence 9, Appli
31	77.5	10.9	9719	4	US-09-700-304-1	Sequence 1, Appli
32	77.5	10.9	11672	4	US-09-441-340-2	Sequence 2, Appli
33	77.5	10.9	15611	4	US-09-441-340-1	Sequence 1, Appli
34	77	10.9	2505	1	US-07-977-434-7	Sequence 7, Appli
35	77	10.9	2505	1	US-08-458-819-7	Sequence 7, Appli
C 36	77	10.9	2505	5	PCT-US91-07035-7	Sequence 7, Appli
C 37	77	10.9	68750	3	US-09-335-409-1	Sequence 1, Appli
C 38	77	10.9	68750	4	US-09-568-102-1	Sequence 1, Appli
C 39	77	10.9	68750	4	US-09-567-969-1	Sequence 1, Appli
C 40	77	10.9	68750	4	US-09-568-480-1	Sequence 1, Appli
C 41	77	10.9	68750	4	US-09-568-486-1	Sequence 1, Appli
C 42	77	10.9	68750	4	US-09-568-472-1	Sequence 1, Appli
C 43	77	10.9	68750	4	US-09-567-899-1	Sequence 1, Appli
C 44	77	10.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 45	77	10.9	4415229	3	US-09-103-840A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-252-991A-15297  
; Sequence 15297, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15297  
; LENGTH: 774  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15297

Alignment Scores:  
Pred. No.: 0.192 Length: 774  
Score: 92.00 Matches: 44  
Percent Similarity: 46.03% Conservative: 14  
Best Local Similarity: 34.92% Mismatches: 51  
Query Match: 12.99% Indels: 17  
DB: 4 Gaps: 5

US-10-087-573-2 (1-141) x US-09-252-991A-15297 (1-774)

QY 25 MetArgGluAlaLeuArgVallySerSerGluArg-----leuAla----- 39  
Db 321 CTGCGGAGAGAGTGTGTCGCGCCCGGAGCTGCGCCAGCTCGCCAGACCGCGGAA 380

```

40 MetLeuArgAlaLeuAlaClyMetCysGlyHisArgValLeuProGlyThrClyAlaSer 59
   ::      |||      |||      |||      |||      |||      |||      |||
381 CTGGAGGCGCTCTCGCGGCTCGCGCGCGCTTCGTTCGCCCGCGCGCGCGCGCAT 440
   |||      |||      |||      |||      |||      |||      |||
60 AlaIle-AlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgPr 79
   |||      |||      |||      |||      |||      |||      |||
441 CGGATCCGCAATCGCAGGTACCCAGCGGCGCAACC-----TGTTCGCTTCGAGGCC 494
   |||      |||      |||      |||      |||      |||      |||
79 o-GlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg-cluMetAsn 98
   |||      |||      |||      |||      |||      |||      |||
495 GACAAGGTACCGACCGCGCGCGCTACGAAGCGGCGCGCGGCTTCGGCCACTG--- 551
   |||      |||      |||      |||      |||      |||      |||
99 LysThrIleSerGlnGluSerAlaArgValAsnHis-----ArgLeuProGluGlyHis 116
   |||      |||      |||      |||      |||      |||      |||
552 -----CTGGAAGAAGCTACCGCGCGCGAGCACAGCGGAGGCGCGCGAAGAGTC 599
   |||      |||      |||      |||      |||      |||      |||
117 ProLeuLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGly 136
   |||      |||      |||      |||      |||      |||      |||
600 GCCTTCAGGCTCTGGTCGTCGGAAACCATCGGCCACCTGGGGCATCGTCGAAAGCCAGGCG 659
   |||      |||      |||      |||      |||      |||      |||
137 ValAsnArgLeu 140
   |||      |||      |||      |||
660 CTCGACGCGCTC 671
   |||      |||      |||      |||

```

```

RESULT 2
US-09-252-991A-15344
; Sequence 15344, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCES: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15344
; LENGTH: 2163
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15344

```

```

Alignment Scores:
Pred. No.:      0.82      Length:      2163
Score:          92.00     Matches:      44
Percent Similarity: 46.03%  Conservative: 14
Best Local Similarity: 34.92% Mismatches:   51
Query Match:      12.99%   Indels:      17
DB:               4       Gaps:        5

US-10-087-573-2 (1-141) x US-09-252-991A-15344 (1-2163)

Qy      25 MetArgAlaLeuLeuArgValLysSerSerGluArg---LeuAla----- 39
      282 CTGCGGAGCAGTTGCTGCGGGCCGGGAACTCGAGCGCGAGCTCCGACCGGCGAA 841
      40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59
      842 CTGAGGCGGTGCTCGCGGCGCTCGCGGCGGTCTGTCGCGCGCGGCGGCGCGCAT 901
      60 AlaIle-AlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgPr 79
      902 CCGATCGCAATCCGCAGGTACCACGCGGCGCAACC-----TGTTCGCTTCGAGGCG 955
      79 o-GlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArg-GluMetAsn 98
      956 GACAAGGTACCGACCGCGCGGCGCTACGAAGCGGCGCGAGGCCCTTCGCCCACTG---- 1012
      99 LysThrIleSerGlnGlnuserAlaIaArgValAsnHis-----ArgLeuProGluGlyHis 116

```

; Sequence 6770, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-07-27  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; LENGTH: 1530  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6770

Alignment Scores:  
Pred. No.: 3-5 Length: 1530  
Score: 84.50 Matches: 43  
Percent Similarity: 41.40% Conservative: 22  
Best Local Similarity: 27.39% Mismatches: 71  
Query Match: 11.94% Indels: 22  
Gaps: 4

US-10-087-573-2 (1-141) x US-09-252-991A-6770 (1-1530)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgPro---ThrPheGly 19  
Db 1130 ATGAACCGGTGAGCCAGACG-----GCGCGGGTGGCGCGATGAGTTTCGT 1083

QY 20 GluThrPhe-----AspValMetArgGluAlaLeuArgValLysSerSerGluArg 37  
Db 1082 GCGATGTTCTGGAAGTCTGTTGGTGGCGGAGACGATGTTCCGCCGCTGTTGGTCGAGAG 1023

QY 38 LeuAla---MetLeuArgAlaLeuAla-----GlyMetCys--- 48  
Db 1022 GTGGCAGCATGATCAACGCCATCGCCACCGCCAGCCCGGGGTGGGATGTCGCCG 963

QY 49 -----GlyHisArgValLeuProGlyThrGlyAlaSerAlaLeuAlaThr 64  
Db 962 ATCAGGCTGACCGGGTGGGAACCGTCTCGCCACCCAGCGAAGCCGAGCGCGGTATC 903

QY 65 ValThrProLysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSer 84  
Db 902 ACCACCGCGGGCGGCGAAGACATGTCAGCGGAGCGCGAAGATCTGCCCGAGT 843

QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGlu 104  
Db 842 AT-CTGGTCTCTTGGCTACGCGCATAGCGGCTGAAAGTCGGGATGTTTCAGCGACAGCGT 784

QY 105 SerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGlu 124  
Db 783 GGCCCGAAGCCGACCATCCCTCAGCCCGCCATGAAGTACCCGGTGAC---GCTGGC 727

QY 125 TyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeuLeu 141  
Db 726 GCCTTCGGCGGCTTCGGCGGCTGGCGCAGCAGTTCGCTGACCGATATGTG 676

RESULT 5  
US-09-252-991A-6969  
; Sequence 6969, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6969  
; LENGTH: 1695  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6969

Alignment Scores:  
Pred. No.: 4-04 Length: 1695  
Score: 84.50 Matches: 43  
Percent Similarity: 41.40% Conservative: 22  
Best Local Similarity: 27.39% Mismatches: 71  
Query Match: 11.94% Indels: 22  
Gaps: 4

US-10-087-573-2 (1-141) x US-09-252-991A-6969 (1-1695)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgPro---ThrPheGly 19  
Db 434 ATGAACCGGTGAGCCAGACG-----GCGCGGGTGGCGCGATGAGTTTCGT 481

QY 20 GluThrPhe-----AspValMetArgGluAlaLeuArgValLysSerSerGluArg 37  
Db 482 GCGATGTTCTGGAAGTCTGTTGGTGGCGGAGACGATGTTCCGCCGCTGTTGGTCGAGAGG 541

QY 38 LeuAla---MetLeuArgAlaLeuAla-----GlyMetCys--- 48  
Db 542 GTGGCAGCATGATCAACGCCATCGCCACCGCCAGCCCGGGGTGGGATGTCGCCG 601

QY 49 -----GlyHisArgValLeuProGlyThrGlyAlaSerAlaLeuAlaThr 64  
Db 602 ATCAGGCTGACCGGGTGGGAACCGTCTCGCCACCCAGCGAAGCCGAGCGCGGTATC 661

QY 65 ValThrProLysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSer 84  
Db 662 ACCACCGCGGGCGGCGAAGACATGTCAGCGGCGCGAAGATCTGCCCGAGT 721

QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGlu 104  
Db 722 AT-CTGGTCTCTTGGCTACGCGCATAGCGGCTGAAAGTCGGGATGTTTCAGCGACAGCGT 780

QY 105 SerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGlu 124  
Db 781 GGCCCGAAGCCGACCATCGCGTCAGCGCGCCATGAAGTACCGGTGAC---GCTGGC 837

QY 125 TyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeuLeu 141  
Db 838 GCCTTCGGCGGCTTCGGCGGCTGGCGCAGCAGTTCGCTGACCGATATGTG 888

RESULT 6  
US-09-252-991A-4734/C  
; Sequence 4734, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4734  
; LENGTH: 1677  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa



US-09-252-991A-4734

Alignment Scores:

Pred. No.: 5.87 Length: 1677  
Score: 83.00 Matches: 35  
Percent Similarity: 35.17% Conservative: 16  
Best Local Similarity: 24.14% Mismatches: 42  
Query Match: 11.72% Indels: 52  
DB: 4 Gaps: 6

US-10-087-573-2 (1-141) x US-09-252-991A-4734 (1-1677)

```
QY 35 SerGluArgLeuAlaMetLeu-----ArgAlaLeuAlaGlyMetCysGly 49
Db 1345 TCCGGCGGCTTCGGGTTCTCCTCAAGAACAAAGGGCGTTCCGCTGCTTCGATGCCGT 1286
QY 50 HisArgValLeuProGlyThr-----56
Db 1285 CATCGACTACCTCCGGCTCCGACCGAGATCCCTGCGATCAAGGGCGTCAGCCCGGACGA 1226
QY 57 -----GlyAlaSerAlaAlaAlaThrValThrProLysGlyAla 70
Db 1225 CGAGACCGTCGAAGACGAGCGTCATGCCGACGACCAACACCGTTCTCGTCCCTGGCGTT 1166
QY 71 SerMetLysLeuLysProPro-----ArgProGlnSerThrLysSer 84
Db 1165 CAAGATCGCTACCGACCGCTTCGTCGGTACCTGACCTTCGCCCGCGTTTACTCCGGCGT 1106
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIle-----101
Db 1105 CCGAGTTCGGCGACTCCGTCCTGAACCTCGGTGAAGAGCAAGAGCGCGTCGCTCG 1046
QY 102 -----SerGlnGluSerAlaArgValAsnHis 110
Db 1045 GATGTGCAGATGCACGCCAACCGAGCGTGAAGAGATCAAGGAAGTGCAGCGGACAT 986
QY 111 ArgLeuPro-----GluGly-----HisProLeuLeuGluLysArg 122
Db 985 CGCTCCCTGATCGGCATGAAGGACGTCAACCGCGTGACACCGCTGTGCTCGATCGAGAA 926
QY 123 AlaGluTyrPheArg 127
Db 925 GCCGATCATCTCGA 911
```

RESULT 7

US-09-252-991A-4774  
; Sequence 4774, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4774  
; LENGTH: 2151  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4774

Alignment Scores:

Pred. No.: 8.35 Length: 2151  
Score: 83.00 Matches: 35  
Percent Similarity: 35.17% Conservative: 16  
Best Local Similarity: 24.14% Mismatches: 42  
Query Match: 11.72% Indels: 52  
DB: 4 Gaps: 6

US-10-087-573-2 (1-141) x US-09-252-991A-4774 (1-2151)

```
QY 35 SerGluArgLeuAlaMetLeu-----ArgAlaLeuAlaGlyMetCysGly 49
Db 816 TCCGGCGGCTTCGGGTTCTCCTCAAGAACAAAGGGCGTTCCGCTGCTTCGATGCCGT 875
QY 50 HisArgValLeuProGlyThr-----56
Db 876 CATCGACTACCTCCGGGTTCCGACCGAGATCCCTGCGATCAAGGGCGTCAGCCCGGACGA 935
QY 57 -----GlyAlaSerAlaAlaAlaThrValThrProLysGlyAla 70
Db 936 CGAGACCGTCGAAGACGAGCGTCATGCCGACGACCAACACCGTTCTCGTCCCTGGCGTT 995
QY 71 SerMetLysLeuLysProPro-----ArgProGlnSerThrLysSer 84
Db 996 CAAGATCGCTACCGACCGCTTCGTCGGTACCTGACCTTCGCCCGCGTTTACTCCGGCGT 1055
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIle-----101
Db 1056 CCGAGTTCGGCGACTCCGTCCTGAACCTCGGTGAAGAGCAAGAGCGCGTCGCTCG 1115
QY 102 -----SerGlnGluSerAlaArgValAsnHis 110
Db 1116 GATGTGCAGATGCACGCCAACCGAGCGTGAAGAGATCAAGGAAGTGCAGCGGACAT 1175
QY 111 ArgLeuPro-----GluGly-----HisProLeuLeuGluLysArg 122
Db 1176 CGCTCCCTGATCGGCATGAAGGACGTCAACCGCGTGACACCGCTGTGCTCGATCGAGAA 1235
QY 123 AlaGluTyrPheArg 127
Db 1236 GCCGATCATCTCGA 1250
```

RESULT 8

US-09-252-991A-12401/c  
; Sequence 12401, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 12401  
; LENGTH: 750  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12401

Alignment Scores:

Pred. No.: 2.14 Length: 750  
Score: 82.50 Matches: 36  
Percent Similarity: 39.71% Conservative: 18  
Best Local Similarity: 26.47% Mismatches: 60  
Query Match: 11.65% Indels: 22  
DB: 4 Gaps: 4

US-10-087-573-2 (1-141) x US-09-252-991A-12401 (1-750)

```
QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
Db 468 GATGGCCCGGAGAGACATGTGTACGCCCGCGCTCCCGCCGCGCCGAGGACCGCCCGC 409
QY 43 AlaLeuAlaGlyMetCysGlyHisArgVal-----LeuProGlyThrGlyAlaSerAla 60
Db: 4
```

Db 408 GGAGCCTCCGCGCAAGCAGGAGGAGCAGAGCGGCGCGCTGGCGCAACTGGTCAAGCA 349  
Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysPro----- 76  
Db 348 GCAGCGACAACAGTCTCGCGCCCTGGCGCGCAGCAGGAGCGGCTGGCGCGGCTCGC 289  
Qy 77 -----ProArgProGlnSerThrLysSer 84  
Db 288 CCGCAGCAGCAGAGAGAACTGGCAGCGCGACAGGAGCGCGCGGCGGCAACTGGACAT 229  
Qy 85 ProGluLeuArgGlu-LeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnG 104  
Db 228 CCGCAGCAGGAACTGCTCAGC-----CTGCAACAGGCGCTGGAACGGCAGCGCGGGA 175  
Qy 104 uSerAlaArgValAsnHisArgLeuProGlu---GlyHisProLeuLeuGluLysArgAl 123  
Db 174 GAACCGCGAAATTGACAGCGCTCTGCTGGAGCAGCGGCGAGCAGTTCACGCGCAACCGCA 115  
Qy 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsn 138  
Db 114 GGAACGACCGGCAACTGGCTTCATCGAGAACCGAGGGCGGCAAC 69

RESULT 9  
US-09-252-991A-12190  
; Sequence 12190, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 12190  
; LENGTH: 1581  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12190

Alignment Scores:  
Pred. No.: 6.15 Length: 1581  
Score: 82.50 Matches: 36  
Percent Similarity: 39.71% Conservative: 18  
Best Local Similarity: 26.47% Mismatches: 60  
Query Match: 11.65% Indels: 22  
DB: 4 Gaps: 4

US-10-087-573-2 (1-141) x US-09-252-991A-12190 (1-1581)

Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 541 GATGCGCGCGAAGAGCAATGTAGCGCGCGGTCCCGCGCGCGCGCGCGCGCGCGCG 600  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgVal-----LeuProGlyThrGlyAlaSerAla 60  
Db 601 GGAGCCTCCGCGCAAGCAGGAGGAGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCG 660  
Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysPro----- 76  
Db 661 GCAGCGACAACAGTCTGGCGCGCGCTGGCGCGGAGCAGGAGCGCGCTGGCGCGGCTCGC 720  
Qy 77 -----ProArgProGlnSerThrLysSer 84  
Db 721 CCGGCGAGCAGAGAGAACTGGCGCGCGCGGAGAACAGGAGCGCGCGCGCGCGCGCG 780  
Qy 85 ProGluLeuArgGlu-LeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnG 104  
Db 781 CCGCGCGCGGAAAGTGCTCAGC-----CTGCAACAGGCGCGCTGGAAACGGCAGCGCGGGA 834

Qy 104 uSerAlaArgValAsnHisArgLeuProGlu---GlyHisProLeuLeuGluLysArgAl 123  
Db 835 GAACCGCGAAATTGACAGCGCGTCTCTGAGCAGCGCGAGCAGTTCACGCGCAACCGCGA 894  
Qy 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsn 138  
Db 895 GGAACGACCGCGCAACTGGCTTCATCGAGAACCGAGGGCGGCGCAAC 940

RESULT 10  
US-09-252-991A-12255  
; Sequence 12255, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 12255  
; LENGTH: 1620  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12255

Alignment Scores:  
Pred. No.: 6.36 Length: 1620  
Score: 82.50 Matches: 36  
Percent Similarity: 39.71% Conservative: 18  
Best Local Similarity: 26.47% Mismatches: 60  
Query Match: 11.65% Indels: 22  
DB: 4 Gaps: 4

US-10-087-573-2 (1-141) x US-09-252-991A-12255 (1-1620)

Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 699 GATGCGCGCGAAGAGCAATGTAGCGCGCGGTCCCGCGCGCGCGCGCGCGCGCG 758  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgVal-----LeuProGlyThrGlyAlaSerAla 60  
Db 759 GGAGCCTCCGCGCAAGCAGCAGGAGGAGCAGCAGCGCGCGCGCGCGCGCGCGCG 818  
Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysPro----- 76  
Db 819 GCAGCGACAACAGTCTGGCGCGCGCTGGCGCGGAGCAGGAGCGCGCTGGCGCGGCTCGC 878  
Qy 77 -----ProArgProGlnSerThrLysSer 84  
Db 879 CCGCAGCAGCAGAGAGAACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 938  
Qy 85 ProGluLeuArgGlu-LeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnG 104  
Db 939 CCGCGCGCGAAGTGTCTCAGC-----CTGCAACAGGCGCGCTGGAAACGGCAGCGCGGGA 992  
Qy 104 uSerAlaArgValAsnHisArgLeuProGlu---GlyHisProLeuLeuGluLysArgAl 123  
Db 993 GAACGCGCAATTGACAGCGCTCTGCTGGAGCAGGCGGAGCAGTTCACGCGCAACCGCGA 1052  
Qy 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsn 138  
Db 1053 GGAACGACCGCGCAACTGGCTTCATCGAGAACCGAGGGCGGCGCAAC 1098

RESULT 11  
US-09-832-441-1  
; Sequence 1, Application US/09832441  
; Patent No. 6506585

```

; GENERAL INFORMATION:
; APPLICANT: Danielsen, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity
; FILE REFERENCE: 10019-200-US
; CURRENT APPLICATION NUMBER: US/09/832,441
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Phaeotrichoconis crotalariae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1797)
US-09-832-441-1

Alignment Scores:
Pred. No.: 10.9 Length: 1797
Score: 81.00 Matches: 27
Percent Similarity: 38.64% Conservative: 7
Best Local Similarity: 30.68% Mismatches: 27
Query Match: 11.44% Indels: 27
DB: Gaps: 3

US-10-087-573-2 (1-141) x US-09-832-441-1 (1-1797)

Qy 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73
Db 1073 CCGCGGTACGGGATGACGGCGGTCCAGACCGGACCCCTTCGGCTCACCTCGCGG 1132
Qy 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89
Db 1133 CTCTGCCACAAACACAAACGACATCCCTTCAAGCCCCCTTCCCGCGCTACCCGTCG 1192
Qy 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109
Db 1193 GCCACGCGACCTTTGGCGG----- 1211
Qy 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121
Db 1212 -CGCGGTCTTCCAAATGGTGGCGGCTACTACACGGGCGGTGGCAGTGGAGGACA 1270
Qy 122 -ArgAlaGluTyrPheArgHis 128
Db 1271 ACGAGCCGACACATCGCCAT 1292

RESULT 13
US-09-252-991A-14835
; Sequence 14835, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14835
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14835

Alignment Scores:
Pred. No.: 10.1 Length: 1422
Score: 80.00 Matches: 36
Percent Similarity: 40.46% Conservative: 17
Best Local Similarity: 27.48% Mismatches: 46
Query Match: 11.30% Indels: 32
DB: Gaps: 7

US-10-087-573-2 (1-141) x US-09-252-991A-14835 (1-1422)

Qy 20 GluThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAla 39
Db 859 GAGGACTATCCGTGTGCGCGCGCTTCTCTACCTGAAGCCCGCGGAGAAACCC 918
Qy 40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59
Db 919 TGGGCCCGAGCCCTGGTA-----CGTTCCGCCAGGCGCGCGCGGCCAG 963
Qy 60 AlaIleAlaThrValThrProLysGly---AlaSerMetLysLeuLys----- 75
Db 964 CGGATC-----GTCACCCGCGGCGGTCTCGCGAGAGATCCAGCGGTGCAG 1014
Qy 76 ---ProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db 76 ---ProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

; GENERAL INFORMATION:
; APPLICANT: Danielsen, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity
; FILE REFERENCE: 10019-200-US
; CURRENT APPLICATION NUMBER: US/09/832,441
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Phaeotrichoconis crotalariae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1797)
US-09-832-441-1

Alignment Scores:
Pred. No.: 10.9 Length: 1797
Score: 81.00 Matches: 27
Percent Similarity: 38.64% Conservative: 7
Best Local Similarity: 30.68% Mismatches: 27
Query Match: 11.44% Indels: 27
DB: Gaps: 3

US-10-087-573-2 (1-141) x US-09-832-441-1 (1-1797)

Qy 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73
Db 1073 CCGCGGTACGGGATGACGGCGGTCCAGACCGGACCCCTTCGGCTCACCTCGCGG 1132
Qy 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89
Db 1133 CTCTGCCACAAACACAAACGACATCCCTTCAAGCCCCCTTCCCGCGCTACCCGTCG 1192
Qy 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValas 109
Db 1193 GCCACGCGACCTTTGGCGG----- 1211
Qy 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121
Db 1212 -CGCGGTCTTCCAAATGGTGGCGGCTACTACACGGGCGGTGGCAGTGGAGGACA 1270
Qy 122 -ArgAlaGluTyrPheArgHis 128
Db 1271 ACGAGCCGACACATCGCCAT 1292

RESULT 12
US-09-833-102-1
; Sequence 1, Application US/09833102
; Patent No. 6521434
; GENERAL INFORMATION:
; APPLICANT: Danielsen, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity
; FILE REFERENCE: 10172-200-US
; CURRENT APPLICATION NUMBER: US/09/833,102
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Phaeotrichoconis crotalariae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1797)
US-09-833-102-1

Alignment Scores:
Pred. No.: 10.9 Length: 1797
Score: 81.00 Matches: 27
```

Db 1015 ATCGCACCGCGCCGAG-----ATGCCCGCGAATACCGCAAGCTCGCGGAACAGCGG 1068  
QY 95 ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGlu 114  
Db 1069 CGCGCGCTGAGC-----GTGAACCTTCGCTTCAGGAA 1101  
QY 115 GlyHisProLeuLeuGluLysArgAla-----GluTyrPhe 126  
Db 1102 GGCAGCGCGCTGCTGCACAAACAGGACCTGCTCGAGCTACAGCGTTTGTCTGACTACCTG 1161  
QY 127 ArgHisLeuArgSerLeuLysSerGlnGlyVal 137  
Db 1162 CGGCAAAATCGGAAGTTGCAGGAGCGCACCGTG 1194  
RESULT 14  
US-09-252-991A-15091  
; Sequence 15091, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15091  
; LENGTH: 1446  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15091  
Alignment Scores:  
Pred. No.: 10.3 Length: 1446  
Score: 80.00 Matches: 36  
Percent Similarity: 40.46% Conservative: 17  
Best Local Similarity: 27.48% Mismatches: 46  
Query Match: 11.30% Indels: 32  
DB: 4 Gaps: 7  
US-10-087-573-2 (1-141) x US-09-252-991A-15091 (1-1446)  
QY 20 GluThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAla 39  
Db 818 GAGGACTATCCGCTGTCGCGCGCCTGTTCTCTACTCGAAGCCGCGGAGAGAACCC 877  
QY 40 MetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer 59  
Db 878 TGGCGCCAGCGCCCTGTA-----CGTTCCGCCAGCGCCCGCGCCAG 922  
QY 60 AlaIleAlaAlaThrValThrProLysGly---AlaSerMetLysLeuLys----- 75  
Db 923 GGATC-----GTCACCGCAGCGCTTCGTCGCGAAGATCCAGCGCGGTGCAG 973  
QY 76 ---ProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 974 ATCGCACCGCGCCGAG-----ATGCCCGCGAATACCGCAAGCTCGCGAACAGCGG 1027  
QY 95 ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGlu 114  
Db 1028 CGCGCGCTGAGC-----GTGAACCTTCGCTTCAGGAA 1060  
QY 115 GlyHisProLeuLeuGluLysArgAla-----GluTyrPhe 126  
Db 1061 GGCAGCGCGCTGCTGCACAAACAGGACCTGCTCGAGTACAGCGTTTGTCTGACTACCTG 1120  
QY 127 ArgHisLeuArgSerLeuLysSerGlnGlyVal 137  
Db 1121 CGGCAAAATCGGAAGTTGCAGGAGCGCACCGTG 1153

RESULT 15  
US-08-660-645A-8/C  
; Sequence 8, Application US/08660645A  
; Patent No. 6087152  
; GENERAL INFORMATION:  
; APPLICANT: Hohmann, Hans-Peter  
; APPLICANT: Fassamontes, Luis  
; APPLICANT: Tessier, Michel  
; APPLICANT: van Loon, Adolphus  
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/660,645A  
; APPLICATION NUMBER: US/08/660,645A  
; FILING DATE: 07-JUN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 95108888.9  
; FILING DATE: 09-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pokras, Bruce A.  
; REGISTRATION NUMBER: 32,748  
; REFERENCE/DOCKET NUMBER: RAN 6002/170  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 235-5801  
; TELEFAX: (201) 235-2363  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1149 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-660-645A-8  
Alignment Scores:  
Pred. No.: 8.51 Length: 1149  
Score: 79.50 Matches: 35  
Percent Similarity: 42.15% Conservative: 16  
Best Local Similarity: 28.93% Mismatches: 50  
Query Match: 11.23% Indels: 20  
DB: 3 Gaps: 4  
US-10-087-573-2 (1-141) x US-08-660-645A-8 (1-1149)  
QY 37 ArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArg----- 51  
Db 590 CGCGTATCTCGATCAGATCGCGGTGGAGCTGAGGGGACAGATAGATGACCGGTAC 531  
QY 52 ---ValLeuProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAla 70  
Db 530 CCGTCCATCTCGCGAACGCGTCCATGATCATCGGGGCGCTCGAGCCATCGGGGCGG 471  
QY 71 SerMetLysLeu-LysProArg-----ProGlnSerThrLysSerProG 86  
Db 470 TCGGTCTCGATCTCGACCGCCACGAAATTCGNAACCCAGGTCAGGTGCGGGGCTCG 411  
QY 86 uLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAl 106  
Db 410 A---CGGCACACCGGGGCTCGATCAGCAGCGACGCTCGATCCGCGAGCGCTCGTCAGC 354

Qy 106 aArgValAsnHisArg-----LeuProGluGlyHisProLe 118  
 Db 353 GTCCGCGCGGTATCGTCCAGCGTGGACATGGTATTCCACCGCAGATCGACACCCCTGC 294  
 Qy 118 uLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAs 138  
 Db 293 AGCAGCCCGATCAGCGCGCGCCGCTCGATCGAGCCATAGCTGTGCTGTCGCGCGCGCGAA 234  
 Qy 138 n 138  
 Db 233 T 233

RESULT 16  
 US-09-298-718-8/c  
 ; Sequence 8, Application US/09298718  
 ; Patent No. 6124113  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hohmann, Hans-Peter  
 ; APPLICANT: Pasamontes, Luis  
 ; APPLICANT: Tessier, Michel  
 ; APPLICANT: van Loon, Adolphus  
 ; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
 ; NUMBER OF SEQUENCES: 47  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hoffmann-La Roche Inc.  
 ; STREET: 340 Kingsland Street  
 ; CITY: Nutley  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 07110  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/298,718  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/660,645  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Pokras, Bruce A.  
 ; REGISTRATION NUMBER: 32,748  
 ; REFERENCE/DOCKET NUMBER: RAN 6002/170  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (201) 235-5801  
 ; TELEFAX: (201) 235-2363  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1149 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-09-298-718-8

Alignment Scores:  
 Pred. No.: 8.51 Length: 1149  
 Score: 79.50 Matches: 35  
 Percent Similarity: 42.15% Conservative: 16  
 Best Local Similarity: 28.93% Mismatches: 50  
 Query Match: 11.23% Indels: 20  
 DB: 3 Gaps: 4

US-10-087-573-2 (1-141) x US-09-298-718-8 (1-1149)  
 Qy 37 ArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArg----- 51  
 Db 590 CGCGTATCTCGATCGATCGCGGTGGAGCTGAAGCGCAGCAGATAGATGAGCGGTAC 531  
 Qy 52 ---ValLeuProGlyThrGlyAlaSerAlaIleAlaIleAlaThrValThrProLysGlyAla 70

Db 530 CGGTCACTCGGAAACCGTCCGCTCATGATCATCGGCGCTCGACGCCATGGGGGGG 471  
 Qy 71 SerMetLysLeu-LysProProArg-----ProGlnSerThrLysSerProG 86  
 Db 470 TCGGTCTCGATCTCGACGCCACCAATTTCTGGAACCCACCGTTCAGGTGCGGGGTCTCG 411  
 Qy 86 uLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAl 106  
 Db 410 A---CGCACACCGCGGTGTCATCAGCAGGAGCTCGATCCGCGAGCGCTCCGCTCAGC 354  
 Qy 106 aArgValAsnHisArg-----LeuProGluGlyHisProLe 118  
 Db 353 GTCCGCGCGGTATCGTCCAGCGTGGACATGGTATTCCACCGCAGATCGACACCCCTGC 294  
 Qy 118 uLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAs 138  
 Db 293 AGCAGCCCGATCAGCGCGCGCCGCTCGATCGAGCCATAGCTGTGCTGTCGCGCGCGAA 234  
 Qy 138 n 138  
 Db 233 T 233

RESULT 17  
 US-09-546-969-8/c  
 ; Sequence 8, Application US/09546969  
 ; Patent No. 6207409  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hohmann, Hans-Peter  
 ; APPLICANT: Pasamontes, Luis  
 ; APPLICANT: Tessier, Michel  
 ; APPLICANT: van Loon, Adolphus  
 ; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
 ; NUMBER OF SEQUENCES: 47  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hoffmann-La Roche Inc.  
 ; STREET: 340 Kingsland Street  
 ; CITY: Nutley  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 07110  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/546,969  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/660,645  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Pokras, Bruce A.  
 ; REGISTRATION NUMBER: 32,748  
 ; REFERENCE/DOCKET NUMBER: RAN 6002/170  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (201) 235-5801  
 ; TELEFAX: (201) 235-2363  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1149 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-09-546-969-8

Alignment Scores:  
 Pred. No.: 8.51 Length: 1149  
 Score: 79.50 Matches: 35  
 Percent Similarity: 42.15% Conservative: 16

```

Best Local Similarity: 28.93% Mismatches: 50
Query Match: 11.23% Indels: 20
DB: 3 Gaps: 4

US-10-087-573-2 (1-141) x US-09-546-969-8 (1-1149)

QY 37 ArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArg----- 51
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
590 CGCGTATCTCGATCAGATCGCGGTGGGACTGAAGGCGAGCAGATAGATGAAGCGGTAC 531
  :||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 52 ---ValLeuProGlyThrGlyAlaSerAlaAlaThrValThrProLysGlyAla 70
  :||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 530 CGGTCCATCTCGGAACGGTCGCTCCATGATCATCGGCGCTCGACGCCATGGGGGGGG 471
  :||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 71 SerMetLysLeu-LysProProArg-----ProGlnSerThrLysSerProcl 86
  |||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 470 TCGGTCTCGATCTCGAGCCGCCACCAATTTCTGNAACCCACGCTCAGGTGCGGGGTCTCG 411
  :||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 86 uLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAl 106
  :||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 410 A---CGGCACCACCGGGGCTGCATCACGACGAGCGCTCGATCCGCGAGCGCTCCGTGACG 354
  :||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 106 aArgValAsnHisArg-----LeuProGluGlyHisProte 118
  |||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 353 GTCGCGCGGTATCGTCCAGCTCGGACATGCGTATTCACCGCAGATCGACACCTGCG 294
  :||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 118 uLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAs 138
  :||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 293 AGCACCCGATCAGCGGCCGCCCTCGATCGAGCCATAGGCTGTCGTGAGGCGGCGGAA 234
  :||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 138 n 138
Db 233 T 233

RESULT 18
US-08-980-832-1
; Sequence 1, Application US/08980832B
; Patent No. 6291204
; GENERAL INFORMATION:
; APPLICANT: Pasamontes, Luis
; APPLICANT: Tsygankov, Yuri
; TITLE OF INVENTION: Improved Fermentative Carotenoid Production
; FILE REFERENCE: Improved Fermentative Carotenoid
; CURRENT APPLICATION NUMBER: US/08/980,832B
; CURRENT FILING DATE: 1997-12-01
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 8625
; TYPE: DNA
; ORGANISM: Flavobacterium sp. R1534
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (8348)..(8349)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (8539)..(8540)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (8581)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (8590)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (8592)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (8602)..(8604)
US-08-980-832-1

Alignment Scores: 147 Length: 8625
Pred. No.:

```

Db 7095 CCGTCCATCTCCGAAACGGTCCGCTCCATGATCATCGGCGCTCGAGCCATCGGGGGCG 7154  
 Qy 71 SerMetLysLeu-LysProProArg-----ProGlnSerThrLysSerProG1 86  
 Db 7155 TCGGTCTCGATCTCGACGCCACCAATTTCTGGAACCCACCGGTGAGTGGGGTCTCG 7214  
 Qy 86 uleuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerA1 106  
 Db 7215 A---CGGCACCACCGGGCTCGATCAGCAGCAGCCCTCGATCCCGGAGCGCTCCGTCAGC 7271  
 Qy 106 aArgValAsnHisArg-----LeuProGluGlyHisProLe 118  
 Db 7272 GTCCGGCCGGTATCGTCCAGCGTCGCGACATGCGTATTCCACCCAGATCGACACCCCTGC 7331  
 Qy 118 uleuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAs 138  
 Db 7332 AGCAGCCCGATCAGCGCGCCCTCGATCAGCCATAGCTGCTCGTACGCGCGCGAA 7391  
 Qy 138 n 138  
 Db 7392 T 7392

RESULT 20  
 US-09-452-239-13/c  
 ; Sequence 13, Application US/09452239  
 ; Patent No. 6465223  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rafalski, Antoni J.  
 ; APPLICANT: Cahn, Rebecca E.  
 ; APPLICANT: Pader, Gary M.  
 ; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase  
 ; FILE REFERENCE: BB1284 US NA  
 ; CURRENT APPLICATION NUMBER: US/09/452,239  
 ; EARLIER FILING DATE: 1999-12-01  
 ; EARLIER FILING DATE: 1998-December-02  
 ; NUMBER OF SEQ ID NOS: 50  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 13  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (483)  
 US-09-452-239-13

Alignment Scores:  
 Pred. No.: 9.01 Length: 997  
 Score: 78.50 Matches: 43  
 Percent Similarity: 40.74% Conservative: 23  
 Best Local Similarity: 26.54% Mismatches: 61  
 Query Match: 11.09% Indels: 35  
 DB: 4 Gaps: 7

US-10-087-573-2 (1-141) x US-09-452-239-13 (1-997)  
 Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgPro----- 16  
 Db 569 TCCGCTCCACGACCGAAGTGAAGCGCGCTCGCGCCCGCGCGCTCCCTCGCGG 510  
 Qy 17 -----ThrPheGlyGluThrPheAspValMetArgGluAlaLeuArgVal 32  
 Db 509 AGCAGCTCGTCCAGCTTCTCCAGCCCTTCCCGCTTGGGAAGTCCACCTTGTGCGCCACC 450  
 Qy 33 -LysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgVa 52  
 Db 449 CCGGCTTTTCCNAGAACCGCC---GCCGATCTCGTAGCTCTCCCTGTCCGGGTGATC 393  
 Qy 52 LeuProGlyThrGlyAlaSerAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 72  
 Db 392 GCCACCACTTCCGCTCCCGGCA---GCGCAGCGCGCTCGCCAGCAGCAGTAGCGG 336

Qy 72 tLysLeuLysProPro-----ArgProGlnSerThrLysSerPro-G 86  
 Db 335 GTGAAGACACCCACCTTCGATTGCTCTTCGCTCGGCCATCTTTCAGCAGCATCCCCAGC 276  
 Qy 86 lueuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerA 106  
 Db 275 AGCTGGCGCTCATCCGCCGACGACTGCATGAACCCACCCACTGGTGTCTTCGCTGATGAG 216  
 Qy 106 laArgValAsnHis-----ArgLeuProGluGlyHisProLeuLeuGluLysArgAla 124  
 Db 215 CGCAGATCCGCGATGACCTCCGGCTCCCGTCCGCGACAC----- 178  
 Qy 124 lufyPheArgHisLeuArgSerLeuLysSerGlnGlyVal-----AsnArgLeu 141  
 Db 177 -----CGTCTGTGCCAGGACATCTTGTACAGGCGCTCGCTCTTGAGCAGCGCTTTC 126  
 Qy 141 le 141  
 Db 125 TT 124

RESULT 21  
 US-09-103-840A-2  
 ; Sequence 2, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; FILE REFERENCE: 24366-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103,840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 4403765  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; FEATURE:  
 ; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 ; OTHER INFORMATION: represent a, t, c or g  
 US-09-103-840A-2

Alignment Scores:  
 Pred. No.: 7.7e+05 Length: 4403765  
 Score: 78.50 Matches: 36  
 Percent Similarity: 40.87% Conservative: 11  
 Best Local Similarity: 31.30% Mismatches: 43  
 Query Match: 11.09% Indels: 25  
 DB: 3 Gaps: 6

US-10-087-573-2 (1-141) x US-09-103-840A-2 (1-4403765)  
 Qy 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
 Db 3542568 ATCCGCGCGCTCGAAGTTCGAGTA-----CGGTAATCCCGGGTTTGTGCGAGTCGCC 3542621  
 Qy 61 lLeAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 72  
 Db 3542622 GTGGCCGATGATGTCAGGTAGATCGCGGGAAGTGGAAATCGCGAGTCAAGAAAGCTTC 3542681  
 Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92  
 Db 3542682 CACCTTCGCCCA-----ACCGTAGGAACCATCCGGCCAGCCAGGACGATTCGCGT 3542735  
 Qy 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArg----- 107  
 Db 3542736 GACCCCGCTCCAGCAGCGCGCTGATGAACGCGCGCAGCGGACGCGGTGATGCC 3542795  
 Qy 108 -----ValAsnHisArgLeuProGluGlyHisProLeuLeu-----Glu 120

Db	3542796	CGGCCGCTCAGCGTGTAGTACCCACCCCTCTCCCGAGCGGAGAACCACTCTCTGCGCTGAT	3542855	...
Qy	121	LysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGln	135	...
Db	3542856	GAGCGGCTCGCCGGTGGTGTAT---CGCGCGCTCGCTAGCGT	3542897	...
RESULT 22				
US-09-103-840A-1				
; Sequence 1, Application US/09103840A				
; Patent No. 6294328				
; GENERAL INFORMATION:				
; APPLICANT: FLEISCHMAN, Robert D.				
; APPLICANT: WHITE, Owen R.				
; APPLICANT: FRASER, Claire M.				
; APPLICANT: VENTER, John C.				
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM				
; TITLE OF INVENTION: TUBERCULOSIS				
; FILE REFERENCE: 24366-20007.00				
; CURRENT APPLICATION NUMBER: US/09/103,840A				
; CURRENT FILING DATE: 1998-06-24				
; NUMBER OF SEQ ID NOS: 2				
; SOFTWARE: PatentIn Ver. 2.1				
; SEQ ID NO 1				
; LENGTH: 4411529				
; TYPE: DNA				
; ORGANISM: Mycobacterium tuberculosis				
; OTHER INFORMATION: H37Rv				
US-09-103-840A-1				
Alignment Scores:				
Pred. No.: 7,71e-05 Length: 4411529				
Score: 78.50 Matches: 36				
Percent Similarity: 40.87% Conservative: 11				
Best Local Similarity: 31.30% Mismatches: 43				
Query Match: 11.09% Indels: 25				
DB: 3 Gaps: 6				
US-10-087-573-2 (1-141) x US-09-103-840A-1 (1-4411529)				
Qy	41	LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla	60	...
Db	3544986	ATCCGGCGCTCGAAGGTTCAGTA-----CGGGTAATCCCGGGTTTGTGCGAGTCGCC	3545039	...
Qy	61	IleAlaAlaThrValThrProLysGlyAla-----SerMet	72	...
Db	3545040	GTGGCCGATGTAGTCCAGGTAGATGCGGGGAAGTGAATCGCGAGCTCAAGAAAGCTTC	3545099	...
Qy	73	LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg	92	...
Db	3545100	CACCTTCGCCCA-----ACCGTAGGAACCATCCGGCCACCCAGGAGCAACGTTCCCGT	3545153	...
Qy	93	LysIleArgGluMetAnLysThrIleSerGlnGluSerAlaArg-----	107	...
Db	3545154	GACCCCGTCCCGAGCAGCGCGCGTATGACGCGCGAGCGCGGAACGTGGTGGTGATGCC	3545213	...
Qy	108	-----ValaenHisArgLeuProGluGlyHisProLeuLeu-----Glu	120	...
Db	3545214	CGCCCGCTCAGCGTCTGTAGTCCACCCCTCTCCCGAGCGGAGAACCACTCTGTGCGTGTAT	3545273	...
Qy	121	LysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGln	135	...
Db	3545274	GAGCGGCTCGCCGGTGGTGTAT---CGCGCGCTCGCTAGCGT	3545315	...
RESULT 23				
US-09-107-532A-1348				
; Sequence 1348, Application US/09107532A				
; Patent No. 6583275				
; GENERAL INFORMATION:				
; APPLICANT: Lynn A Doucette-Stamm and David Bush				
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO				
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS				
; NUMBER OF SEQUENCES: 7310				



```

; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3856 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: HXB2
; US-07-743-357-20

Alignment Scores:
Pred. No.: 79.1 Length: 3856
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44
Query Match: 10.95% Indels: 35
DB: 2 Gaps: 6

US-10-087-573-2 (1-141) x US-07-743-357-20 (1-3856)

QY 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
DB 3049 TTTGACTGTTTTCAGACTCTGCTATAAGA----- 3078
QY 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52
DB 3079 AAGGCTTTATTAGACACATAGTTCAGCTAGGTGTGAATATCAAGCAGACATACAAAG 3138
QY 53 LeuProGlyThrGlyAlaSerAlaIleAlaIleAlaThrValThrProLysGlyAlaSerMet 72
DB 3139 GTAGGATCTCTACATACTTGGCACTAGCAGCATTAATAACACCAAAA----- 3186
QY 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92
DB 3187 AAGATAAGCCACTTTCCTAGTGTAGTAAGAACTGACAGAGATAGA----- 3234
QY 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112
DB 3235 -----TGGAAACAGCCCAAGAGACCAAGGGCCACAGAGGAGCCACACAATG 3282
QY 113 ProGluGlyHis--ProLeuLeuGluLys--ArgAlaGluTyPheArgHisLeuArgS 131

```

```

Db 3283 ---AATGACACTAGAGCTTTTAGAGAGAGCTTAAGAAATGAAGCTGTTAGACATTTTCCTA 3339
QY 131 erLeuLysSerGlnGlyValAsnArgLeuIle 141
DB 3340 GGATTTGGCTCCATGGCTTAGGCGCAACATATC 3371

RESULT 25
US-07-743-357-21
; Sequence 21, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: HXB2
; US-07-743-357-21

Alignment Scores:
Pred. No.: 86.7 Length: 4113
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44
Query Match: 10.95% Indels: 35
DB: 2 Gaps: 6

US-10-087-573-2 (1-141) x US-07-743-357-21 (1-4113)

QY 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
DB 3306 TTTGACTGTTTTCAGACTCTGCTATAAGA----- 3335
QY 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52
DB 3336 AAGGCTTTATTAGACACATAGTTCAGCTAGGTGTGAATATCAAGCAGACATACAAAG 3395

```



```
; LOCATION: 113..1648
; OTHER INFORMATION: /product= "gag"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1408..4452
; OTHER INFORMATION: /product= "pol"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4367..4975
; OTHER INFORMATION: /product= "sor"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 5560..8148
; OTHER INFORMATION: /product= "env"
US-08-463-210-4

Alignment Scores:
Pred. No.: 260 Length: 8933
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44
Query Match: 10.95% Indels: 35
DB: 3 Gaps: 6

US-10-087-573-2 (1-141) x US-08-463-210-4 (1-8933)
Qy 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
Db 4733 TTTCAGTCTTTTCAGACTCTCTATAGA-----MetCysGlyHisArgVal 52
Qy 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52
Db 4763 AAGGCCCTATTAGGACACATAGTTAGCCCTAGGTGTAATATCAAGCAGGACATAACAAG 4822
Qy 53 LeuProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMet 72
Db 4823 GTAGGATCTCTACATACATTGCGACTAGCAGCATTAATAACACCAAAA-----4870
Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92
Db 4871 AAGATAAAGCCACCTTTGCTAGTGTAGAAACTGACAGAGGATAGA-----4918
Qy 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112
Db 4919 -----TGGACAGCCCCAGAGACCAAGGCCAGAGGCCAGGACCAATG 4966
Qy 113 ProGluGlyHis--ProLeuLeuGluLys---ArgAlaGluTyrPheArgHisLeuArg 131
Db 4967 ---AATGGACACATAGAGCTTTTAGAGGAGCTTTAGAGATGAAGCTGTTAGACATTTCCTA 5023
Qy 131 erLeuLysSerGlnGlyValAsnArgLeu 141
Db 5024 GGATTGGCTCCATGGCTTAGGGCAACATATC 5055

RESULT 28
US-09-620-958A-3
; Sequence 3, Application US/09620958A
; Patent No. 6294338
; GENERAL INFORMATION:
; APPLICANT: Nunomura, Kiyotada
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
; FILE REFERENCE: GP104-02.JT
; CURRENT APPLICATION NUMBER: US/09/620,958A
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 8933
; TYPE: RNA
; ORGANISM: Human Immunodeficiency Virus
; FEATURE:
; OTHER INFORMATION: Sequence of the IAC-Asrc pseudo target
; Patent No. 6294338
; NAME/KEY: mutation
; LOCATION: (4135)...(4155)
; OTHER INFORMATION: Mutated positions: 4135, 4140-1, 4145, 4150,
; OTHER INFORMATION: 4152-3, 4155
US-09-620-958A-4

Alignment Scores:
Pred. No.: 260 Length: 8933
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44
```

```
; OTHER INFORMATION: Sequence of transcripts produced from the BH10
; Patent No. 6294338
; OTHER INFORMATION: plasmid.
US-09-620-958A-3

Alignment Scores:
Pred. No.: 260 Length: 8933
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44
Query Match: 10.95% Indels: 35
DB: 3 Gaps: 6

US-10-087-573-2 (1-141) x US-09-620-958A-3 (1-8933)
Qy 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
Db 4733 UUUGACUUUUUUCAGACUCUCUAUAGA-----MetCysGlyHisArgVal 52
Qy 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52
Db 4763 AAGCCUUUUUAGGACACAUAGUUUAGCCCUUAGGUGUGUAUAUUAUUAUUAUUAUUA 4822
Qy 53 LeuProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMet 72
Db 4823 GUAGGAUUCUACAAUACUUGGCACUAGUAGCCCUUAGGUGUGUAUAUUAUUAUUAUUA 4870
Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92
Db 4871 AAGAUAAAGCCACCUUUGCUAGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4918
Qy 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112
Db 4919 -----UGGACAAAGCCCCAGAGACCAAGGCCAGGCGGACGACCAAAU 4966
Qy 113 ProGluGlyHis--ProLeuLeuGluLys---ArgAlaGluTyrPheArgHisLeuArg 131
Db 4967 ---AUGGACACUAGAGCUUUUAGAGGCUUAAGAGCUUAAGAGCUGUUAUUAUUAUUA 5023
Qy 131 erLeuLysSerGlnGlyValAsnArgLeu 141
Db 5024 GGAUUUGCUUCCAUUGCUUAGGCGCAACAUUC 5055

RESULT 29
US-09-620-958A-4
; Sequence 4, Application US/09620958A
; Patent No. 6294338
; GENERAL INFORMATION:
; APPLICANT: Nunomura, Kiyotada
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD
; FILE REFERENCE: GP104-02.JT
; CURRENT APPLICATION NUMBER: US/09/620,958A
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 8933
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Sequence of the IAC-Asrc pseudo target
; Patent No. 6294338
; NAME/KEY: mutation
; LOCATION: (4135)...(4155)
; OTHER INFORMATION: Mutated positions: 4135, 4140-1, 4145, 4150,
; OTHER INFORMATION: 4152-3, 4155
US-09-620-958A-4

Alignment Scores:
Pred. No.: 260 Length: 8933
Score: 77.50 Matches: 37
Percent Similarity: 40.15% Conservative: 16
Best Local Similarity: 28.03% Mismatches: 44
```

Query Match: 10.95% Indels: 35  
DB: 3 Gaps: 6

US-10-087-573-2 (1-141) x US-09-620-958A-4 (1-8933)

Qy 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41  
Db 4733 UUGACUGUUUUCAGACUCUCGUAAGA-----MetCysGlyHisArgVal 52  
Qy 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52  
Db 4763 AAGGCCUUUAGGACACAUAGUAGCCUAGGUGUGUAUACGAGGACGACAUACAAG 4822  
Qy 53 LeuProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMet 72  
Db 4823 GUAGGAUCUCUACAAUACUUGGCACUAGCAGCAUUAUAACACCAAAA-----4870  
Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92  
Db 4871 AAGAUAAGCCACCCUUGCCUAGUCUUAAGAAACUGACAGGAUAGA-----4918  
Qy 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112  
Db 4919 -----UGGAACAAGCCCCAGAACAGCCAGAGGGCCACAGAGGACGACACAAG 4966  
Qy 113 ProGluGlyHis--ProLeuLeuGluLys---ArgAlaGluTyrPheArgHisLeuArgS 131  
Db 4967 ---AUGGACACUAGAGCUUUAAGAGGAGCUUAAGAUAAGAGCUUUAAGCAUUAUCCUA 5023  
Qy 131 erLeuLysSerGlnGlyValAsnArgLeuIle 141  
Db 5024 GGAUUGGCCUCCAUUGGCUUAGGGCAACAUAUC 5055

RESULT 30

US-09-620-958A-9  
; Sequence 9, Application US/09620958A  
; Patent No. 6294338  
; GENERAL INFORMATION:  
; APPLICANT: Nunomura, Kiyotada  
; TITLE OF INVENTION: POLYNUCLEOTIDE AMPLIFICATION METHOD  
; FILE REFERENCE: GP104-02 UT  
; CURRENT APPLICATION NUMBER: US/09/620,958A  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9  
; LENGTH: 8933  
; TYPE: RNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Sequence of the IAC-Bscr pseudo target  
; Patent No. 6294338  
; NAME/KEY: mutation  
; LOCATION: (4140)...(4159)  
; OTHER INFORMATION: Mutated positions: 4140-42, 4145-47, 4152,  
; OTHER INFORMATION: 4156-57, 4159

US-09-620-958A-9

Alignment Scores:  
Pred. No.: 260 Length: 8933  
Score: 77.50 Matches: 37  
Percent Similarity: 40.15% Conservative: 16  
Best Local Similarity: 28.03% Mismatches: 44  
Query Match: 10.95% Indels: 35  
DB: 3 Gaps: 6

US-10-087-573-2 (1-141) x US-09-620-958A-9 (1-8933)

Qy 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41  
Db 4733 UUGACUGUUUUCAGACUCUCGUAAGA-----MetCysGlyHisArgVal 52  
Qy 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52

Db 4763 AAGGCCUUUAGGACACAUAGUAGCCUAGGUGUGUAUACAAGCAGGACGACAUACAAG 4822

Qy 53 LeuProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMet 72

Db 4823 GUAGGAUCUCUACAAUACUUGGCACUAGCAGCAUUAUAACACCAAAA-----4870

Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92

Db 4871 AAGAUAAGCCACCCUUGCCUAGUCUUAAGAAACUGACAGGAUAGA-----4918

Qy 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112

Db 4919 -----UGGAACAAGCCCCAGAACAGCCAGAGGGCCACAGAGGACGACACAAG 4966

Qy 113 ProGluGlyHis--ProLeuLeuGluLys---ArgAlaGluTyrPheArgHisLeuArgS 131

Db 4967 ---AUGGACACUAGAGCUUUAAGAGGAGCUUAAGAUAAGAGCUUUAAGCAUUAUCCUA 5023

Qy 131 erLeuLysSerGlnGlyValAsnArgLeuIle 141

Db 5024 GGAUUGGCCUCCAUUGGCUUAGGGCAACAUAUC 5055

RESULT 31

US-09-700-304-1  
; Sequence 1, Application US/09700304  
; Patent No. 650623  
; GENERAL INFORMATION:  
; APPLICANT: GeneCure, LLC  
; TITLE OF INVENTION: Replication Defective HIV Vaccine  
; FILE REFERENCE: 00714PCTUS  
; CURRENT APPLICATION NUMBER: US/09/700,304  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: PCT/US99/10523  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/085,115  
; PRIOR FILING DATE: 1998-05-12  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 9719  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1

US-09-700-304-1

Alignment Scores:  
Pred. No.: 293 Length: 9719  
Score: 77.50 Matches: 37  
Percent Similarity: 40.15% Conservative: 16  
Best Local Similarity: 28.03% Mismatches: 44  
Query Match: 10.95% Indels: 35  
DB: 4 Gaps: 6

US-10-087-573-2 (1-141) x US-09-700-304-1 (1-9719)

Qy 22 PheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeu 41

Db 5374 TTTGACTGTTTTTCAGACTCTGCTATAAGA-----MetCysGlyHisArgVal 52

Qy 42 ArgAlaLeuAlaGly-----MetCysGlyHisArgVal 52

Db 5404 AAGGCCUATTAGGACACATAGTTAGCCCTAGGTGTAATATCAAGCAGGACATACAAG 5463

Qy 53 LeuProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMet 72

Db 5464 GTAGGATCTCTACAATACTTGGCAGCTAGCAGCATTAATAACACCAAAA-----5511

Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92

Db 5512 AAGATAAAGCCACCTTTGCTCTAGTTACCAAACTGACAGAGGATAGA-----5559

Qy 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112

```

Db      5560 -----TGGAAACAGCCCAAGAGACCAAGGGCCACAGAGGGAGGACCAACAATG 5607
Qy      113 ProGluGlyHis--ProLeuLeuGluLys--ArgAlaGluTyrPheArgHisLeuArgS 131
Db      5608 ----AATGGACACTAGAGCTTTTAGAGAGCTTAAGAACTGTTAGACATTTTCCTA 5664
Qy      131 erLeuLysSerGlnGlyValAsnArgLeu141
Db      5665 GGATTTGGCTCCATGCTTAGGGCAACATATC 5696

```

RESULT 32

```

US-09-441-340-2
; Sequence 2, Application US/09441340
; Patent No. 6448476
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; TITLE OF INVENTION: Phosphonate Metabolizing Plants
; FILE REFERENCE: 38-21(15303)
; CURRENT APPLICATION NUMBER: US/09/441,340
; EARLIER FILING DATE: 1999-11-16
; EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 11672
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-441-340-2

```

```

Alignment Scores:
Pred. No.:      379      Length:      11672
Score:          77.50    Matches:      35
Percent Similarity: 39.72% Conservative: 21
Best Local Similarity: 24.82% Mismatches: 50
Query Match:     10.95% Indels:      30
DB:              4      Gaps:        4

```

US-10-087-573-2 (1-141) x US-09-441-340-2 (1-11672)

```

Qy      7 ThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg 26
Db      8536 ACCACTCGCGGGCCAGCGCCAGTTCGCCAACCGCGAGAGATTCGGGAATATTATCAGG 8595
Qy      27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAla 43
Db      8596 GCAAACTACTCCCTCACTGATCGCGAGATCGAGCAGTACGAAGAAGAGACAACTGGCGCTCG 8655
Qy      44 LeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAla 63
Db      8656 -----CCGCACGCTGTCGACGCCGAATTCGCGAAT 8685
Qy      64 ThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLys 83
Db      8686 CCATCGCGCCCTGTGCGCGCGGCGGAAAATTGGCTTGCAGCCAGATGACGCCACCC 8745
Qy      84 SerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr-IleSerG1 103
Db      8746 ACGCCCACTTG-----CTGAATCTCACCAGCTTGCA 8778
Qy      103 nGluSerAlaArgValAsnHisArgLeuProGluGly-----Hi 116
Db      8779 GCGTGATCGCGGAATTTCCACACAGTTCGAAGCGCGGGAAGCTCGCGCAAGCATGGCA 8838
Qy      116 sProLeuLeuGlySerGlnAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnG1 136
Db      8839 TGAACGTGCTGATGGCGCGCGGAATATTGTGCGCGCGGCTCGCACTCCGCAACGTGG 8898
Qy      136 Y 136
Db      8899 C 8899

```

```

RESULT 33
US-09-441-340-1
; Sequence 1, Application US/09441340
; Patent No. 6448476
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; TITLE OF INVENTION: Phosphonate Metabolizing Plants
; FILE REFERENCE: 38-21(15303)
; CURRENT APPLICATION NUMBER: US/09/441,340
; EARLIER FILING DATE: 1999-11-16
; EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 15611
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-441-340-1

```

```

Alignment Scores:
Pred. No.:      572      Length:      15611
Score:          77.50    Matches:      35
Percent Similarity: 39.72% Conservative: 21
Best Local Similarity: 24.82% Mismatches: 55
Query Match:     10.95% Indels:      30
DB:              4      Gaps:        4

```

US-10-087-573-2 (1-141) x US-09-441-340-1 (1-15611)

```

Qy      7 ThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg 26
Db      12772 ACCACTCGCGGGCCAGCGCCAGTTCGCCAACCGCGAGAGATTCGCGAATATTATCAGG 12831
Qy      27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAla 43
Db      12832 GCAAACTACTCCCTCACTGATCGCGAGATCGAGCAGTACGAAGAAGAGCAACTGGCGCTCG 12891
Qy      44 LeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAla 63
Db      12892 -----CCGCACGCTGTCGCGCAATTCGCGAAT 12921
Qy      64 ThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLys 83
Db      12922 CCATCGCGCCCTGTGCGCGCGGCGGAAAATTGGCTTGCAGCCAGATGACGCCACCC 12981
Qy      84 SerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr-IleSerG1 103
Db      12982 ACGCCCACTTG-----CTGAATCTCACCAGCTTGCA 13014
Qy      103 nGluSerAlaArgValAsnHisArgLeuProGluGly-----Hi 116
Db      13015 GCGTGATCGCGGAATTTCCACACAGTTCGAAGCGCGGGAAGCTCGCGCAAGCATGGCA 13074
Qy      116 sProLeuLeuGlySerGlnAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnG1 136
Db      13075 TGAACGTGCTGATGGCGCGCGGAATATTGTGCGCGCGGCTCGCACTCCGCAACGTGG 13134
Qy      136 Y 136
Db      13135 C 13135

```

RESULT 34

```

US-07-977-434-7
; Sequence 7, Application US/07977434
; Patent No. 5466591
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abramson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:

```



QY	1	MetGluSerThrSerThrThrAenPheValAlaGluAuaAsnArgProThrPheGlyGlu	20
Db	1144	CTGAGCCCTCCAAACACACCCCGGGGGTGGCC-----CGGCGCTACGGGGGGAG	1197
QY	21	---ThrPheAspValMetArgGluAlaLeuLeu-----	30
Db	1198	TGGACGGAGGACGCCGCCACCGCGGCTCTCGCCGAGCGGCTCCAGCAAACTCTTTG	1257
QY	31	---ArgVallysserSerGluArgLeuAlaMetLeu-----	41
Db	1258	GAACGCTTCAAGGAGAGAAAGCTCTTTGCTCTTACCAAGAGGTGGAAAGCCCTC	1317
QY	42	---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer	59
Db	1318	TCCCGGGTCTTGCCACATGGA-GGCCACCGGGTAAGGCTGGAGTGCCTATCTAAA	1376
QY	60	-----AlaIleAlaAlaThrValThrProLysGlyAlaSer-----	72
Db	1377	GGCCTTTCCTGGAGCTTCGGAGGAGATTCCCGCTCGAGAGAGAGTCTTCGCGCT	1436
QY	73	-----LysLeuysProProArgProGlnSerThrLysSerProGluLeu-ArgG	89
Db	1437	GGCGGGCCACCCCTTCAACCTGAACCTCCCTGACCAAGTACAGCGGGTGCCTTTGACGA	1496
QY	89	uLeuSerArgLysIleArgGluMetAsnLysThrIleSer-----GlnGluSe	105
Db	1497	GCTT-----AGGCTTCCCGCTGGGCAAGACGCAAAAGACGGGAGCGCTCCACCA	1550
QY	105	rAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluysArgAlaGlu	125
Db	1551	CGCGCGGTGCTGGAGGCGCTCAGGAGGCGCCACCCCATCGTGAGAAGATCCTCCAGCA	1610
QY	125	rPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu	140
Db	1611	C---CGGAGGCTCACAAGCTCAAGACACCTACGTGGACCCCTC	1653
RESULT 36			
PCT-US91-07035-7			
; Sequence 7, Application PC/TUS9107035			
; GENERAL INFORMATION:			
; APPLICANT: Gelfand, David H.			
; APPLICANT: Abramson, Richard D.			
; TITLE OF INVENTION: 5' TO 3', EXONUCLEASE MUTATIONS OF			
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES			
; NUMBER OF SEQUENCES: 38			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Cetus Corporation			
; STREET: 1400 Fifty-third Street			
; CITY: Emeryville			
; STATE: California			
; ZIP: 94608			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Wordperfect 5.0			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: PCT/US91/07035			
; FILING DATE: 19910930			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 590,490			
; FILING DATE: 28-SEP-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 590,466			
; FILING DATE: 28-SEP-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 590,213			
; FILING DATE: 28-SEP-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 523,394			
; FILING DATE: 15-MAY-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 523,394			
; FILING DATE: 15-MAY-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 523,394			
; FILING DATE: 15-MAY-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 523,394			
; FILING DATE: 15-MAY-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 523,394			
; FILING DATE: 15-MAY-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 523,394			
; FILING DATE: 15-MAY-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 523,394			
; FILING DATE: 15-MAY-1990			

	Alignment Scores:			
Pred. No.:	48.9	Length:	2505	
Score:	77.00	Matches:	53	
Percent Similarity:	41.48%	Conservative:	20	
Best Local Similarity:	30.11%	Mismatches:	62	
Query Match:	10.88%	Indels:	42	
DB:	1	Gaps:	9	
		US-10-087-573-2 (1-141) x	US-08-458-819-7 (1-2505)	

[illegible][illegible]



[illegible]

## Alignment Scores:

Pred. No.:	5.3e+03	Length:	68750
Score:	77.00	Matches:	34
Percent Similarity:	45.19%	Conservative:	13
Best Local Similarity:	32.62%	Mismatches:	28
Query Match:	10.88%	Indels:	29
DB:	4	Gaps:	5

US-10-087-573-2 (1-141) x US-09-568-480-1 (1-68750)

Qy	8	ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArgGlu	27
Db	54914	ACCAACGATCTCGCGAGCTGGCCAGCTTTTCGGCGAG-----CAACGC	54870
Qy	28	AlaLeuLeuArgValLysSerSerGluArgLeu-----	38
Db	54869	GTCCCTCTGCTCGTCCGTCATGCCCGCAGAGCCTCGAGATCTCGGGCATCGTTCTCGAA	54810
Qy	39	AlaMetLeu-ArgAlaLeuAlaGly-----MetCysGlyHisArgVal--	52
Db	54809	GCTCTCTCCCGCTCGGTGGCGGAGCGTGGTCCGCCGGCATTCCGGAACAGAAATGTC	54750
Qy	53	-----LeuProGlyThrGlyAla-----SerAlaIleAlaAlaTh	64
Db	54749	TAGCAAGCTCCGGCTCAGAGCTGCTACGTTAGGGTAGGTCCATAGCAGGGTCCGCCGCAC	54690
Qy	64	rValThrProLysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSe	84
Db	54689	GGTGATGCCGAGCGGCTCGATCGGTTGGGAGCTCCAGGCCCTATCAGCGAGTCCAT	54630
Qy	84	rProGluLeu	87
Db	54629	GCCGAGATTG	54620

Search completed: November 17, 2003, 14:38:38

Job time : 1606 secs



```
Alignment Scores:
Pred. No.: 2,04e-77 Length: 1135
Score: 708.00 Matches: 141
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-087-573-2 (1-141) x US-10-087-573-1 (1-1135)
Qy 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20
Db 75 ATGGAGTCGACATCAACACCAACCAACTTTGTCCGAGAACCGTCCACCTTTGGTGAG 134
Qy 21 ThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMet 40
Db 135 ACGTTTGATGTGATGAGGGAAGCTTTGCTTCGTGTAAAGTCTCTGAAACGCTTGGCAATG 194
Qy 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60
Db 195 CTCAGAGCGCTTGCAGGAATGTCCGCTCACCGCTCTCTCTGCGACTGGTCTTCTGCG 254
Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGln 80
Db 255 ATAGCGGCAACGGTAACCCCAAGGGGCTTCGATGAAGCTTAAACCAACCGCTCCGAG 314
Qy 81 SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr 100
Db 315 TCAACGAAAGTCTCCGAGCTCAGGAGCTCAGGAGCTGTACGGAAGATTCCGGAATGAATAAGACT 374
Qy 101 IleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGlu 120
Db 375 ATAAGTCAGGAATCAGCTCGGTAAACCAACCGGTTCGCGGAAGCCACCTCTCTTAGAG 434
Qy 121 LysArgAlaGluTyPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140
Db 435 AAGCGGGCAGAGATATTTTCGTACCTTAGATCTCTTAAAGCCAGGAGTCAATAGACTC 494
Qy 141 Ile 141
Db 495 ATC 497

RESULT 2
US-10-087-573-3
; Sequence 3, Application US/10087573
; Publication No. US20030165872A1
; GENERAL INFORMATION:
; APPLICANT: SCHEITERS, Theodor P
; APPLICANT: CARCY, Bernard PD
; APPLICANT: DRACULOVSKI, Pascal R
; APPLICANT: GORENFLOT, Andre F
; TITLE OF INVENTION: BABESIA CANIS VACINE
; FILE REFERENCE: SCHEITERS
; CURRENT APPLICATION NUMBER: US/10/087,573
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: EP 01200816.5
; PRIOR FILING DATE: 2001-03-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Babesia canis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75) .. (929)
US-10-087-573-3
Alignment Scores:
Pred. No.: 5,67e-74 Length: 1134
Score: 680.00 Matches: 140
Percent Similarity: 99.29% Conservative: 0

Best Local Similarity: 99.29% Mismatches: 1
Query Match: 96.05% Indels: 1
DB: 12 Gaps: 0

US-10-087-573-2 (1-141) x US-10-087-573-3 (1-1134)
Qy 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20
Db 75 ATGGAGTCGACATCAACACCAACCAACTTTGTCCGAGAACCGTCCACCTTTGGTGAG 134
Qy 21 ThrPheAspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMet 40
Db 135 ACGTTTGATGTGATGAGGGAAGCTTTGCTTCGTGTAAAGTCTCTGAAACGCTTGGCAATG 194
Qy 41 LeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60
Db 195 CTCAGAGCGCTTGCAGGAATGTCCGCTCACCGCTCTCTCTGCGACTGGTCTTCTGCG 254
Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArgProGln 80
Db 255 ATAGCGGCAACGGTAACCCCAAGGGGCTTCGATGAAGCTTAAACCAACCGCTCCGAG 314
Qy 81 SerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThr 100
Db 315 TCAACGAAAGTCTCCGAGCTCAGGAGCTCAGGAGCTGTACGGAAGATTCCGGAATGAATAAGACT 374
Qy 101 IleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGlu 120
Db 375 ATAAGTCAGGAATCAGCTCGGTAAACCAACCGGTTCGCGGAAGCCACCTCTCTTAGAG 434
Qy 121 LysArgAlaGluTyPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140
Db 435 AAGCGGGCAGAGATATTTTCGTACCTTAGATCTCTTAAAGCCAGGAGTCAATAGACTC 493
Qy 141 Ile 141
Db 494 ATC 496

RESULT 3
US-10-160-758-10/c
; Sequence 10, Application US/10160758
; Publication No. US20030036076A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-089C
; CURRENT APPLICATION NUMBER: US/10/160,758
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 10531
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-160-758-10
Alignment Scores:
Pred. No.: 2.1 Length: 10531
Score: 92.50 Matches: 38
Percent Similarity: 46.58% Conservative: 30
Best Local Similarity: 26.03% Mismatches: 52
Query Match: 13.06% Indels: 26
DB: 14 Gaps: 6

US-10-087-573-2 (1-141) x US-10-160-758-10 (1-10531)
Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
```

```

; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 667
; LENGTH: 1659
; TYPE: DNA
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-796-692-667

Alignment Scores:
Pred. No.: 0.369 Length: 1659
Score: 90.00 Matches: 40
Percent Similarity: 41.54% Conservative: 14
Best Local Similarity: 30.77% Mismatches: 41
Query Match: 12.71% Indels: 35
DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x US-09-796-692-667 (1-1659)
Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26
Db 501 ACACGCTTGGTGGCCCTCCAGGGTGCAGCGGGCTCCCGAGGGCGAGCATCAGGCC 442
Qy 27 -----GluAlaLeuArgValLysSerSerGluArgGluAlaMetLeuArg 42
Db 441 CCATCGGGATCTCCACACCATGACTGTGCGAGTCTCGAGCCCTCGCAATGACGTTG 382
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
Db 381 TCATTG-----TGCGGGCACAGCGGATGTCTAGCACAGGGGCTGTGTGGCCACAG 331
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
Db 330 ACCGTGGGGCATTTCTTCCACAGCTCCAGTCTTGCCAGGGCGAGCCAGGAGGCC 271
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db 270 CCTCCCCCG-----CTGGCCCTCAGATC 247
Qy 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108
Db 246 AGGGCCACAACTTAGGGTTGACAGACAGAGAACCACTGTCCAGGTGTCTGTGAGAGC 187

```

```

; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 667
; LENGTH: 1659
; TYPE: DNA
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-796-692-667

Alignment Scores:
Pred. No.: 0.369 Length: 1659
Score: 90.00 Matches: 40
Percent Similarity: 41.54% Conservative: 14
Best Local Similarity: 30.77% Mismatches: 41
Query Match: 12.71% Indels: 35
DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x US-09-796-692-667 (1-1659)
Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26
Db 501 ACACGCTTGGTGGCCCTCCAGGGTGCAGCGGGCTCCCGAGGGCGAGCATCAGGCC 442
Qy 27 -----GluAlaLeuArgValLysSerSerGluArgGluAlaMetLeuArg 42
Db 441 CCATCGGGATCTCCACACCATGACTGTGCGAGTCTCGAGCCCTCGCAATGACGTTG 382
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
Db 381 TCATTG-----TGCGGGCACAGCGGATGTCTAGCACAGGGGCTGTGTGGCCACAG 331
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
Db 330 ACCGTGGGGCATTTCTTCCACAGCTCCAGTCTTGCCAGGGCGAGCCAGGAGGCC 271
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db 270 CCTCCCCCG-----CTGGCCCTCAGATC 247
Qy 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108
Db 246 AGGGCCACAACTTAGGGTTGACAGACAGAGAACCACTGTCCAGGTGTCTGTGAGAGC 187

```

```

; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 667
; LENGTH: 1659
; TYPE: DNA
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-796-692-667

Alignment Scores:
Pred. No.: 0.369 Length: 1659
Score: 90.00 Matches: 40
Percent Similarity: 41.54% Conservative: 14
Best Local Similarity: 30.77% Mismatches: 41
Query Match: 12.71% Indels: 35
DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x US-09-796-692-667 (1-1659)
Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26
Db 501 ACACGCTTGGTGGCCCTCCAGGGTGCAGCGGGCTCCCGAGGGCGAGCATCAGGCC 442
Qy 27 -----GluAlaLeuArgValLysSerSerGluArgGluAlaMetLeuArg 42
Db 441 CCATCGGGATCTCCACACCATGACTGTGCGAGTCTCGAGCCCTCGCAATGACGTTG 382
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
Db 381 TCATTG-----TGCGGGCACAGCGGATGTCTAGCACAGGGGCTGTGTGGCCACAG 331
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
Db 330 ACCGTGGGGCATTTCTTCCACAGCTCCAGTCTTGCCAGGGCGAGCCAGGAGGCC 271
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db 270 CCTCCCCCG-----CTGGCCCTCAGATC 247
Qy 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108
Db 246 AGGGCCACAACTTAGGGTTGACAGACAGAGAACCACTGTCCAGGTGTCTGTGAGAGC 187

```

```

; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 667
; LENGTH: 1659
; TYPE: DNA
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-796-692-667

Alignment Scores:
Pred. No.: 0.369 Length: 1659
Score: 90.00 Matches: 40
Percent Similarity: 41.54% Conservative: 14
Best Local Similarity: 30.77% Mismatches: 41
Query Match: 12.71% Indels: 35
DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x US-09-796-692-667 (1-1659)
Qy 8 ThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26
Db 501 ACACGCTTGGTGGCCCTCCAGGGTGCAGCGGGCTCCCGAGGGCGAGCATCAGGCC 442
Qy 27 -----GluAlaLeuArgValLysSerSerGluArgGluAlaMetLeuArg 42
Db 441 CCATCGGGATCTCCACACCATGACTGTGCGAGTCTCGAGCCCTCGCAATGACGTTG 382
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
Db 381 TCATTG-----TGCGGGCACAGCGGATGTCTAGCACAGGGGCTGTGTGGCCACAG 331
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
Db 330 ACCGTGGGGCATTTCTTCCACAGCTCCAGTCTTGCCAGGGCGAGCCAGGAGGCC 271
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db 270 CCTCCCCCG-----CTGGCCCTCAGATC 247
Qy 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108
Db 246 AGGGCCACAACTTAGGGTTGACAGACAGAGAACCACTGTCCAGGTGTCTGTGAGAGC 187

```

QY 108 laenHisArgLeuProGluGlyHisPro 117  
Db 186 CGCACATCTTCATAGCACTGGTCGGCCT 159

RESULT 6  
US-10-040-862-667/c  
; Sequence 667, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-01352035  
; CURRENT APPLICATION NUMBER: US/10/040,862  
; CURRENT FILING DATE: 2001-11-06  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: US 09/796,692  
; PRIOR FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 10467  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 667  
; LENGTH: 1659  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-040-862-667

Alignment Scores:  
Pred. No.: 0.369 Length: 1659  
Score: 90.00 Matches: 40  
Percent Similarity: 41.54% Conservative: 14  
Best Local Similarity: 30.77% Mismatches: 41  
Query Match: 12.71% Indels: 35  
DB: 14 Gaps: 6

US-10-087-573-2 (1-141) x US-10-040-862-667 (1-1659)

QY 8 ThrAspPheValAlaGluAsnArgProThrPheGlyGluThrPheAspValMetArg--- 26  
Db 501 ACACGTTGGTGGCCCTCCAGGGTGACAGCGGGCTCCCGAGGGGCAGCATCAGGCC 442

QY 27 -----GluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 441 CCATCCGGGATCTCCACACCATGACTGTCAGTCTCCGAGCCACTGGCAATGACGTTG 382

QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 381 TCATTG-----TGCGGGCACAGGCGATGCTAGCACAGGGGCTGTGTGCCACAG 331  
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 330 ACCGTGGCGCATCTTGTCCACACGTCCTAGTCTGCCAGGGGCGACACAGGAAGGCC 271  
QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 270 CTTCCCGCG-----CTGGGCTCACAGATC 247  
QY 95 ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaIleAla 108  
Db 246 AGGCCACAAACTTAGGGTTGACAGCACAGACCACTGCCAGGTGTGTGTGAGACG 187  
QY 108 laenHisArgLeuProGluGlyHisPro 117  
Db 186 CGCACATCTTCATAGCACTGGTCGGCCT 159

RESULT 7  
US-09-918-995-36271/c  
; Sequence 36271, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 36271  
; LENGTH: 432  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-36271

Alignment Scores:  
Pred. No.: 0.0952 Length: 432  
Score: 88.50 Matches: 34  
Percent Similarity: 44.66% Conservative: 12  
Best Local Similarity: 33.01% Mismatches: 27  
Query Match: 12.50% Indels: 30  
DB: 11 Gaps: 5

US-10-087-573-2 (1-141) x US-09-918-995-36271 (1-432)

QY 30 LeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGly 49  
Db 428 ATGACTGTGCAGTCTCTCGAGCCACTGGCAATGACGTTGTATTG-----TGCGGG 378

QY 50 HisArgValLeuProGlyThrGlyAla-----SerAlaIleAlaAla 63  
Db 377 CACGAGCGATGCTAGCACAGGGGCTGTGTGGCCACAGACCGTGGCGGCATTCTTTGCC 318

QY 64 Thr-----ValThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSer 81  
Db 317 ACAGTCCAGTCTTGGCCAGGGGCGACACAGGAGGCCCTCCCGG----- 270

QY 82 ThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn----- 98  
Db 269 -----CTGGCCTCAGATCAGGGCCACAAACTTAGGGTTG 234

QY 99 -----LysThrIleSerGlnGlu-SerAlaIleValAsnHisArgLeuProGluG 115  
Db 233 ACAGCACAGAGCCACTGTCTCCAGGTGGTCTGTGAGAGCGGCACATCTTCATAGCACTGG 174

QY 115 yHisPro 117  
Db 173 TCGGCCT 167

RESULT 8
US-09-918-995-36013/c
; Sequence 36013, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36013
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-36013
Alignment Scores:
Pred. No.: 0.0996 Length: 447
Score: 88.50 Matches: 34
Percent Similarity: 44.66% Conservative: 12
Best Local Similarity: 33.01% Mismatches: 27
Query Match: 12.50% Indels: 20
DB: 11 Gaps: 5
US-10-087-573-2 (1-141) x US-09-918-995-36013 (1-447)
Qy 30 LeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGly 49
Db 389 ATGATGTCGAGTCCTCGAGGCATCGCATGAGCTTGTCTATG-----TGGGGG 339
Qy 50 HisArgValLeuProGlyThrGlyAla-----SerAlaIleAlaAla 63
Db 338 CACCAGGCGATCTCTAGCACAGGGGCTGTGTGGCCACAGCGTGGGGCGCATTTGTGCC 279
Qy 64 Thr-----ValThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSer 81
Db 278 ACCTGTCAGTCTTCCAGGGGAGCAGCAGGAGGGCCCTCCCTCC----- 231
Qy 82 ThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn----- 98
Db 230 -----CTGGCCTCACAGATCAGGGCCACAACTTAGGGTTG 195
Qy 99 -----LysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGluGl 115
Db 194 TCAGCACAGACCCACTGTCAGGAGTGGTCTGTGAGACGGCCACATCTTCATAGCACTGG 135
Qy 115 yHisPro 117
Db 134 TCGGCT 128
RESULT 9
US-09-918-995-8937
; Sequence 8937, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8937
; LENGTH: 486
US-10-087-573-2 (1-141) x US-09-918-995-8937 (1-486)
Qy 20 GluThrPheAspValMetArgGluAlaLeuLeuArg-----Val 32
Db 160 GATCGTTTAAAGTTATGAAGCAAGAAAGAAATAGGAACAGAAATATCAAGATGAAGTT 219
Qy 33 LysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgVal 52
Db 220 AAGAAGCTTCAAAAGGCAGCTCTCTCTGAAAGCT-----TCTCCCAAAAAA 264
Qy 53 LeuProGlyThr---GlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSer 71
Db 265 GCACCTGGTACTAAAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 324
Qy 72 MetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSer 91
Db 325 GTTCAGCAAAAAGATCACCGCGCGAGTAAAGAGGCTCCAGCCCATAGGTTCTCTGCC 384
Qy 92 ArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArg 111
Db 385 AGAAGGCACAGGCCATATAAGCAGCGCTCTCTCAAAAGCTCAGAGGGTCAAAAAGCTC 444
Qy 112 LeuProGluGlyHisProLysLeuGluLysArgAlaGluTyrPheArgHisLeu 129
Db 445 CAGCCCAAGAAAGCAC---CTGCTCCAA-----AGGCATCTG 477
RESULT 10
US-09-864-636A-89
; Sequence 89, Application US/09864636A
; Publication No. US20030104378A1
; GENERAL INFORMATION:
; APPLICANT: Third Wave Technologies
; APPLICANT: Allwal, Hatim
; APPLICANT: Bartholomay, Christian
; APPLICANT: Chehak, LuAnne
; TITLE OF INVENTION: Detection of RNA Sequences
; FILE REFERENCE: FORS-04944
; CURRENT APPLICATION NUMBER: US/09/864,636A
; CURRENT FILING DATE: 2002-10-15
; NUMBER OF SEQ ID NOS: 2840
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 89
; LENGTH: 2517
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-864-636A-89
Alignment Scores:
Pred. No.: 3.51 Length: 2517
Score: 84.00 Matches: 55
Percent Similarity: 41.48% Conservative: 18
Best Local Similarity: 31.25% Mismatches: 62
Query Match: 11.86% Indels: 42
DB: 11 Gaps: 9

```

RESULT 11
US-09-758-282-120
; Sequence 120, Application US/09758282
; Publication No. US2003013439A1
; GENERAL INFORMATION:
; APPLICANT: Ma, Wu-Po
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamicheva, Natalie E.
; APPLICANT: Allawi, Hatim T.
; APPLICANT: Schaefer, James J.
; APPLICANT: Neri, Bruce P.
; TITLE OF INVENTION: Improved Enzymes for the Detection of Specific Nucleic
; FILE REFERENCE: FORS-04323
; CURRENT APPLICATION NUMBER: US/09/758,282
; CURRENT FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 120
; LENGTH: 2517
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-758-282-120

Alignment Scores:
Pred. No.: 3.51
Score: 84.00
Percent Similarity: 41.48%
Best Local Similarity: 31.25%
Query Match: 11.86%
DB: 12
Length: 2517
Matches: 55
Conservative: 18
Mismatches: 62
Indels: 42
Gaps: 9

```



; SEQ ID NO 89  
; LENGTH: 2517  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-10-084-839-89  
Alignment Scores:  
Pred. No.: 3.51 Length: 2517  
Score: 84.00 Matches: 55  
Percent Similarity: 41.48% Conservative: 18  
Best Local Similarity: 31.25% Mismatches: 62  
Query Match: 11.86% Indels: 42  
DB: 12 Gaps: 9  
US-10-087-573-2 (1-141) x US-10-084-839-89 (1-2517)  
Qy 1 MetGluSerThrSerThrThrThrAsnPhValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1141 CTGACCTTCGACACACCCCGGAGGGGTGGCG-----CGCGCTACGGGGGGAG 1194  
Qy 21 --ThrPheAspValMetArgGluAlaLeu----- 30  
Db 1195 TGGACGGAGGACGGCGCCCGGAGGCTCTCTCGGAGAGGCTCCATCGGAACCTCCTT 1254  
Qy 31 --ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1255 AAGCGCTCGAGGGGAGGAGAACTCTTTGGCTCTACCAAGGAGGTGGAAAGCCCTC 1314  
Qy 42 --ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSer--- 59  
Db 1315 TCCCGGTCTCGGCCCATATGGA--GGCACGGGGGTGCGCTGACCTGACCTATCTCAG 1373  
Qy 60 -----AlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeu 74  
Db 1374 GGCCTTGCTGCTGAGCTTGGCGAGGAGATCGCCGCTCGAGGCGAGGT-----CTT 1427  
Qy 75 LysProArgPro-----GlnSerThrLysSerProGlu 86  
Db 1428 CCGCTGCGCGGCCACCCCTTCAACCTCACTCCCGGAGCAGCGCAAGCGCTCCACAG 1547  
Qy 87 LeuArgGluLeuSerArgLysIleArgGlu--MetAsnLysThrIleSerGlnGlu---Se 105  
Db 1488 TGACGAGTAGGCTTCCCGCATCGGACAGGAGAGCGGCAAGCGCTCCACAG 1547  
Qy 105 xAlaArgValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTy 125  
Db 1548 CGCGCGCTCTCGAGGCCCTCCCGAGGCCACCCCATCGTGAGAGATCTCTGCAGTA 1607  
Qy 125 rPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1608 C---CGGAGCTCACAAGCTGAAGACCTACATTGACCCCTTG 1650  
RESULT 13  
US-10-029-386-5420  
; Sequence 5420, Application US/10029386  
; Publication No. US2003019704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 5420  
; LENGTH: 505  
; TYPE: DNA  
; ORGANISM: Homo sapiens

; FEATURE:  
; OTHER INFORMATION: MAP TO AC012331.9  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1  
; OTHER INFORMATION: EST HUMAN HIT: B1906112.1, EVALUE 3.00e-31  
; OTHER INFORMATION: NT HIT: 273645.1, EVALUE 0.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: P01714, EVALUE 6.00e-30  
US-10-029-386-5420  
Alignment Scores:  
Pred. No.: 0.64 Length: 505  
Score: 82.50 Matches: 31  
Percent Similarity: 43.48% Conservative: 9  
Best Local Similarity: 33.70% Mismatches: 41  
Query Match: 11.65% Indels: 11  
DB: 12 Gaps: 4  
US-10-087-573-2 (1-141) x US-10-029-386-5420 (1-505)  
Qy 35 SerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuPro 54  
Db 224 AGCCAGAGATCGCTCAGGATCCTTGAGGGCGGTCTACTGTATCATAGA---TGACCA 280  
Qy 55 GlyThrGlyAlaSer---AlaIleAlaAlaThrValThrProLysGlyAlaSerMetLys 73  
Db 281 GCACGGGGCGCTCGCTCTCTGTGTACCAAGTGTTCATAAGAGCCTTCCATGCTGT 340  
Qy 74 LeuLys-----ProProArgProGlnSerThrLysSerProGlu 86  
Db 341 CTCCTGGCAGGTGATCTCGGCCATTTGTCCCAAGGCCACAGACACTGCGAGCCACTGAG 400  
Qy 87 LeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAla 106  
Db 401 TTGGCCAG-----AGGAATAACGGAGCCCAAGAGGAGGAGGAGGTGAGTCTG 454  
Qy 107 ArgValAsnHisArgLeuProGluGlyHisProLeu 118  
Db 455 AGGATGAGGGCCTCATTCACAGAACCCACACCCCTG 490  
RESULT 14  
US-10-156-761-7288  
; Sequence 7288, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: HORIKAWA, JUN  
; APPLICANT: ISHIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: RATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7288  
; LENGTH: 747  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(747)  
US-10-156-761-7288  
Alignment Scores:

Pred. No.: 1.08 Length: 747  
Score: 82.50 Matches: 30  
Percent Similarity: 40.66% Conservative: 7  
Best Local Similarity: 32.97% Mismatches: 34  
Query Match: 11.65% Indels: 20  
DB: 14 Gaps: 3  
US-10-087-573-2 (1-141) x US-10-156-761-7288 (1-747)

QY 58 AlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLysLeuLysProPro 77  
Db 473 GCATCGCTCGCGCGCTGTGTGGCGCTGCGCCGAGAGGGCTTCGGTGGCTGCACCTCGCGGAG 532  
QY 78 ArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMet 97  
Db 533 CGGCAAGATCAACCACTCAAGCTGCTGCCGAGGGCGCGCCGAGCGGAGACCGG 592  
QY 98 AsnLysThr-----IleSerGlnGluSer-AlaAr 107  
Db 593 TCCTCGACATGTGTGACCCAGATGACGAGAGGGCTTCGGTGGCTGCACCTCGCGGAG 652  
QY 107 gValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyPheAr 127  
Db 653 AGTGCGCCACCGCTGCGCGAGGGATCC-----GCTGGTCTCCATCACCG 700  
QY 127 gHisLeuArgSerLeuLysSerGlnGlyVal 137  
Db 701 GCAT-----GAACAAGGAGTG 716

RESULT 15

US-10-156-761-1

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; PRIOR FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 1

; LENGTH: 9025608

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4187715)

; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Alignment Scores:

Pred. No.: 2.63e+05 Length: 9025608  
Score: 82.50 Matches: 30  
Percent Similarity: 40.66% Conservative: 7  
Best Local Similarity: 32.97% Mismatches: 34  
Query Match: 11.65% Indels: 20  
DB: 14 Gaps: 3

US-10-087-573-2 (1-141) x US-10-156-761-1 (1-9025608)

QY 58 AlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLysLeuLysProPro 77

Db 8718299 GCATCGCTCGCGCGCTGTGTGGCGCTGCGCCGAGAGGGCTTCGGTGGCTGCACCT 8718359

QY 78 ArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMet 97  
Db 8718359 CGGCAAGATCAACCACTCAAGCTGCTGCCGAGGGCGCGCCGAGCGGAGACCGG 8718418  
QY 98 AsnLysThr-----IleSerGlnGluSer-AlaAr 107  
Db 8718419 TCCTCGACATGTGTGACCCAGATGACGAGAGGGCTTCGGTGGCTGCACCTCGCGGAG 8718478  
QY 107 gValAsnHisArgLeuProGluGlyHisProLeuLeuGluLysArgAlaGluTyPheAr 127  
Db 8718479 AGTGCGCCACCGCTGCGCGAGGGATCC-----GCTGGTCTCCATCACCG 8718526  
QY 127 gHisLeuArgSerLeuLysSerGlnGlyVal 137  
Db 8718527 GCAT-----GAACAAGGAGTG 8718542

RESULT 16

US-10-156-761-3722

; Sequence 3722, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; PRIOR FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 3722

; LENGTH: 786

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(786)

US-10-156-761-3722

Alignment Scores:

Pred. No.: 1.33 Length: 786  
Score: 82.00 Matches: 47  
Percent Similarity: 35.98% Conservative: 21  
Best Local Similarity: 24.87% Mismatches: 58  
Query Match: 11.58% Indels: 64  
DB: 14 Gaps: 8

US-10-087-573-2 (1-141) x US-10-156-761-3722 (1-786)

QY 4 ThrSerThrThrThrAsnPheValAlaGluAsnArgProThr-----Phe 18

Db 4 ACCTCACTCCGCGACGCTGCTTGGCGACCTGGACCCCGAGCGATCGTGGCGTGTTC 63

QY 19 GlyGluThrPheAspValMetArgGluAlaLeu-----LeuArgValLysSerSerGlu 36

Db 64 CGGCAGATCGCGACCGACGCTGCGCGAGGCCATCGACCGTGGCGGATTCAAGAGGGCGGAA 123

QY 37 ArgLeu-----AlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgVal 52

Db 124 AAGCTCCCTCGGAAGCTGAGCTCGT-CGAGGACTACGGAGTGTCCCGGATGACCGTCCG 182

QY 53 -----LeuProGlyThrGlyAlaSerAlaIleAlaIle----- 63

Db 183 GAACCTCTTCTCCATCTCTCCAGGCGGAGGGCTCGTCCATCGCGGAGCACCGAAGAGCGGT 242

QY	65	valThrProGlyGlyAlaSerMetIysLeuLysPro-----ProArg 78
Db	6000095	GTGATCCCCCTACTCCGCGAGCTTGCCCTTGGCGCGGTGGGGCGGTGAGGACCAACGG 6000036
QY	79	ProGInSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn 98
Db	6000035	CCCTCTCTCGTCAGCGCCACCGAGTGTCCAGTCGAGGACCAAGTGCCTCGGTGGTGTG 5999976
QY	99	LysThrIleSerGlnGluSerAlaArgValAlaHisArgLeuProGluGlyHisProLeu 118
Db	5999975	ATGACGGTCAGTCGCTCTCGAGCAC-CTCGGCTCTTCGTTGTCGCGAGGACACCATCGG 5999917
QY	119	LeuGluLysArgAlaGluTyrrPheargHisLeuArgSerLeuLysSerGlnGlyValAla 138
Db	5999916	CTCGATCGCCAGCAGAGACCGCGGCACCAAGCTTCGACCTTGCCTCCGCGCGGTCTCCAC 5999857
RESULT 18		
US-08-319-974A-1		
; Sequence 1, Application US/08319974A		
; Publication No. US20030104576A1		
; GENERAL INFORMATION:		
; APPLICANT: Nyce, Jonathan W.		
; TITLE OF INVENTION: Attenuated Viruses and Method of Making		
; NUMBER OF SEQUENCES: 1		
; CORRESPONDENCE ADDRESS:		
; ADDRESSER: Kenneth D. Sibley		
; STREET: Post Office Box 34009		
; CITY: Charlotte		
; STATE: No. US20030104576A1th Carolina		
; COUNTRY: USA		
; ZIP: 28234		
; COMPUTER READABLE FORM:		
; MEDIUM TYPE: Floppy disk		
; COMPUTER: IBM PC compatible		
; OPERATING SYSTEM: PC-DOS/MS-DOS		
; SOFTWARE: Patent In Release #1.0, Version #1.30		
; CURRENT APPLICATION DATA:		
; APPLICATION NUMBER: US/08/319,974A		
; FILING DATE: 07-OCT-1994		
; CLASSIFICATION: 435		
; ATTORNEY/AGENT INFORMATION:		
; NAME: Sibley, Kenneth D.		
; REGISTRATION NUMBER: 31,665		
; REFERENCE/DOCKET NUMBER: 5218-27		
; TELECOMMUNICATION INFORMATION:		
; TELEPHONE: 919-881-3140		
; TELEFAX: 919-881-3175		
; TELEX: 575102		
; INFORMATION FOR SEQ ID NO: 1:		
; SEQUENCE CHARACTERISTICS:		
; LENGTH: 9718 base pairs		
; TYPE: nucleic acid		
; STRANDEDNESS: single		
; TOPOLOGY: linear		
; MOLECULE TYPE: cDNA		
US-08-319-974A-1		
Alignment Scores:		
Pred. No.: 42.6 Length: 9718		
Score: 81.50 Matches: 34		
Percent Similarity: 37.50% Conservative: 11		
Best Local Similarity: 28.33% Mismatches: 38		
Query Match: 11.51% Indels: 37		
DB: 8 Gaps: 6		
US-10-087-573-2 (1-141) x US-08-319-974A-1 (1-9718)		
QY	22	PheAspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeu 41
Db	5373	TTTCGACTGTTTTCGACTCGCGATACGC-----MetCysGlyHisArgVal 52

Db 5403 AAGCGCTTACTCGGACACATCGTTTCGCGCGCTCGAATATCAACGGGACATAACAAAG 5462  
Qy 53 LeuProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMet 72  
Db 5463 TCGCGATCGCTCAATACCTCGCGCTCGCGGCTTAATAACCGCGAAA----- 5510  
Qy 73 LysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArg 92  
Db 5511 AAGATAAAGCCGCGTTCGCTCGGTAGCAAACTGACGAGATCGA----- 5558  
Qy 93 LysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHisArgLeu 112  
Db 5559 -----TGGAAACAGCCCAAGAGACCAAGGGCCACAGAGGGAGCCACAAATG 5606  
Qy 113 ProGluGlyHis-----ProLeuLeuGluLysArgAlaGluTyrPheArg 127  
Db 5607 ---AATGGACACTAGAGCTTTAGAGAGCTTAAGAACAGCGGTTCGCCATTTTCGCC 5663

RESULT 19

US-09-833-102-1  
; Sequence 1, Application US/09833102  
; Patent No. US2002006652A1  
; GENERAL INFORMATION:  
; APPLICANT: Danielson, Steffen  
; APPLICANT: Schneider, Palle  
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity  
; FILE REFERENCE: 10172.200-US  
; CURRENT APPLICATION NUMBER: US/09/833,102  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 1  
; LENGTH: 1797  
; TYPE: DNA  
; ORGANISM: Phaeotrichoconis crotalariae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1797)  
US-09-833-102-1

Alignment Scores:  
Pred. No.: 5.26 Length: 1797  
Score: 81.00 Matches: 27  
Percent Similarity: 38.64% Conservatve: 7  
Best Local Similarity: 30.68% Mismatches: 27  
Query Match: 11.44% Indels: 27  
DB: 9 Gaps: 3

US-10-087-573-2 (1-141) x US-09-833-102-1 (1-1797)

Qy 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73  
Db 1073 CCGCGGTACGGATGACGGCGGTCACAGACCGGACCCCTCTGGCTCACCTCGCGG 1132  
Qy 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89  
Db 1133 CTCTGCGCAACAAACAGACATCCCTTCAAGCCCGCTTCCCGCGCTACCGGTCCG 1192  
Qy 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAs 109  
Db 1193 GCCACGGACCTTTGGCGG----- 1211  
Qy 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121  
Db 1212 -CGCGGTCTTCCAAATGGTGGCGCGCTACTACAACGGCGGCGTGGGACGAGACA 1270  
Qy 122 -ArgAlaGluTyrPheArgHis 128  
Db 1271 ACGAGCCGACACATCGCCAT 1292

RESULT 20

US-09-832-441-1

; Sequence 1, Application US/09832441  
; Patent No. US2002009434A1  
; GENERAL INFORMATION:  
; APPLICANT: Danielson, Steffen  
; APPLICANT: Schneider, Palle  
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity  
; FILE REFERENCE: 10019.200-US  
; CURRENT APPLICATION NUMBER: US/09/832,441  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 1  
; LENGTH: 1797  
; TYPE: DNA  
; ORGANISM: Phaeotrichoconis crotalariae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1797)  
US-09-832-441-1

Alignment Scores:  
Pred. No.: 5.26 Length: 1797  
Score: 81.00 Matches: 27  
Percent Similarity: 38.64% Conservatve: 7  
Best Local Similarity: 30.68% Mismatches: 27  
Query Match: 11.44% Indels: 27  
DB: 9 Gaps: 3

US-10-087-573-2 (1-141) x US-09-832-441-1 (1-1797)

Qy 54 ProGlyThrGlyAlaSerAlaIleAlaThrValThrProLysGlyAlaSerMetLys 73  
Db 1073 CCGCGGTACGGATGACGGCGGTCACAGACCGGACCCCTCTGGCTCACCTCGCGG 1132  
Qy 74 LeuLysPro-----ProArgProGlnSerThrLysSerProGluLeuArgGlu 89  
Db 1133 CTCTGCGCAACAAACAGACATCCCTTCAAGCCCGCTTCCCGCGCTACCGGTCCG 1192  
Qy 90 LeuSerArgLysIle-ArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAs 109  
Db 1193 GCCACGGACCTTTGGCGG----- 1211  
Qy 109 nHisArgLeuProGluGlyHisProLeuLeuGluLys----- 121  
Db 1212 -CGCGGTCTTCCAAATGGTGGCGCGCTACTACAACGGCGGCGTGGGACGAGACA 1270  
Qy 122 -ArgAlaGluTyrPheArgHis 128  
Db 1271 ACGAGCCGACACATCGCCAT 1292

RESULT 21

US-10-156-761-2584/c  
; Sequence 2584, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAVOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2584  
; LENGTH: 1902



```
/ LENGTH: 3530
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-969-708-93

Alignment Scores:
Pred. No.: 12.8 Length: 3530
Score: 81.00 Matches: 30
Percent Similarity: 38.04% Conservative: 5
Best Local Similarity: 32.61% Mismatches: 32
Query Match: 11.44% Indels: 25
DB: 10 Gaps: 2

US-10-087-573-2 (1-141) x US-09-969-708-93 (1-3530)
Qy 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50
Db 595 CGCACGCTCAGCTCTCCAAAGTTGGCTTCCCGGAGGATATGCGAGCATTA 536
Qy 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaThrVa 65
Db 535 CCGAGTTGTATCATACACGTGGCGGCGGCGGCTCTCTCGCGGAGCATGTT 476
Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85
Db 535 CCGAGTTGTATCATACACGTGGCGGCGGCGGCTCTCTCGCGGAGCATGTT 476
Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85
Db 475 GCGCGCCGTCGGAGCCCGGAGCCCGGAGCCCGGAGCCCGGAGCCCGGAGCC 416
Qy 85 oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe 105
Db 415 CCAG----- 412

RESULT 24
US-09-969-708-622/c
/ Sequence 622, Application US/09969708
/ Patent No. US20020102532A1
/ GENERAL INFORMATION:
/ APPLICANT: Augustus, Meena
/ TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
/ FILE REFERENCE: 689290-70
/ CURRENT APPLICATION NUMBER: US/09/969,708
/ PRIOR FILING DATE: 2001-10-03
/ PRIOR APPLICATION NUMBER: US/60/237,606
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: US/60/237,608
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: US/60/237,425
/ PRIOR FILING DATE: 2000-10-03
/ NUMBER OF SEQ ID NOS: 658
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 622
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-969-708-622

Alignment Scores:
Pred. No.: 12.8 Length: 3530
Score: 81.00 Matches: 30
Percent Similarity: 38.04% Conservative: 5
Best Local Similarity: 32.61% Mismatches: 32
Query Match: 11.44% Indels: 25
DB: 10 Gaps: 2

US-10-087-573-2 (1-141) x US-09-969-708-622 (1-3530)
Qy 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50
Db 595 CGCACGCTCAGCTCTCCAAAGTTGGCTTCCCGGAGGATATGCGAGCATTA 536
```

```
Qy 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaThrVa 65
Db 535 CCGAGTTGTATCATACACGTGGCGGCGGCGGCTCTCTCGCGGAGCATGTT 476
Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85
Db 475 GCGCGCCGTCGGAGCCCGGAGCCCGGAGCCCGGAGCCCGGAGCCCGGAGCC 416
Qy 85 oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe 105
Db 415 CCAG----- 412

Qy 105 rAlaArgValAsnHisArgLeuProGluGlyHis 116
Db 411 -AGACGGGCTGGGCGGCGGCGGAAACCCGAGGGCCAC 379

RESULT 25
US-09-880-107-3771/c
/ Sequence 3771, Application US/09880107
/ Patent No. US20020142981A1
/ GENERAL INFORMATION:
/ APPLICANT: Horne, Darci T.
/ APPLICANT: Vockley, Joseph G.
/ APPLICANT: Scherf, Uwe
/ APPLICANT: Gene Logic, Inc.
/ TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
/ FILE REFERENCE: 44921-5028-WO
/ CURRENT APPLICATION NUMBER: US/09/880,107
/ CURRENT FILING DATE: 2001-06-14
/ PRIOR FILING DATE: 2000-06-14
/ PRIOR APPLICATION NUMBER: US 60/211,379
/ PRIOR FILING DATE: 2000-06-14
/ PRIOR APPLICATION NUMBER: US 60/237,054
/ PRIOR FILING DATE: 2000-10-02
/ NUMBER OF SEQ ID NOS: 3950
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 3771
/ LENGTH: 3530
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: Genbank Accession No. US20020142981A1 X67247
US-09-880-107-3771

Alignment Scores:
Pred. No.: 12.8 Length: 3530
Score: 81.00 Matches: 30
Percent Similarity: 38.04% Conservative: 5
Best Local Similarity: 32.61% Mismatches: 32
Query Match: 11.44% Indels: 25
DB: 10 Gaps: 2

US-10-087-573-2 (1-141) x US-09-880-107-3771 (1-3530)
Qy 31 ArgValLysSerSerGluArgLeuAlaMet-LeuArgAlaLeuAlaGlyMetCysGlyHi 50
Db 595 CGCACGCTCAGCTCTCCAAAGTTGGCTTCCCGGAGGATATGCGAGCATTA 536
Qy 50 sArgValLeu-----ProGlyThrGlyAlaSerAlaAlaThrVa 65
Db 535 CCGAGTTGTATCATACACGTGGCGGCGGCGGCTCTCTCGCGGAGCATGTT 476
Qy 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSerThrLysSerPr 85
Db 475 GCGCGCCGTCGGAGCCCGGAGCCCGGAGCCCGGAGCCCGGAGCCCGGAGCC 416
Qy 85 oGluLeuArgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSe 105
Db 415 CCAG----- 412

Qy 105 rAlaArgValAsnHisArgLeuProGluGlyHis 116
Db 411 -AGACGGGCTGGGCGGCGGCGGAAACCCGAGGGCCAC 379
```

RESULT 26  
US-10-156-761-1190/c  
; Sequence 1190, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156.761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1190  
; LENGTH: 993  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(993)  
US-10-156-761-1190  
Alignment Scores:  
Pred. No.: 2,76 Length: 993  
Score: 80.50 Matches: 30  
Percent Similarity: 48.84% Conservative: 12  
Best Local Similarity: 34.88% Mismatches: 41  
Query Match: 11.37% Indels: 4  
DB: 14 Gaps: 1  
US-10-087-573-2 (1-141) x US-10-156-761-1190 (1-993)  
QY 30 LeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGly 49  
Db 783 CTTCGCCGTATGCTCGGACCGCCAGCGAGCGGTTTCAGGCCGCGCTCGCGGC 724  
QY 50 HisArgValLeuProGlyThrGly---AlaSerAlaIleAlaAlaThrVal--ThrProL 68  
Db 723 GAACGCGGGTCCGGGGACCGCGATCCCGATGGCTTCCCGCGGCTTGACACCG 664  
QY 68 ySGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSerProGluLeuA 88  
Db 663 CAGGGACTTCTCG-AGATTCTACACCGAGGTGTCCAGCACCGAGTCTAGCCGGTCA 605  
QY 88 rgGluLeuSerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgV 108  
Db 604 GGAGCTGCTGAAGTCTGGTGGGTGAATCGATCACCGTGTCCGCGCGAGCGCGCA 545  
QY 108 alaAsnHisArgLeu 112  
Db 544 CGAAGTCCGCGTTG 531  
RESULT 27  
US-09-864-636A-96  
; Sequence 96, Application US/09864636A  
; Publication No. US20030104378A1  
; GENERAL INFORMATION:  
; APPLICANT: Third Wave Technologies  
; APPLICANT: Allwai, Hatim  
; APPLICANT: Bartholomay, Christian  
; APPLICANT: Chehak, LuAnne  
; TITLE OF INVENTION: Detection of RNA Sequences  
; FILE REFERENCE: FORS-04944  
; CURRENT APPLICATION NUMBER: US/09/864.636A

; CURRENT FILING DATE: 2002-10-15  
; NUMBER OF SEQ ID NOS: 2640  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 96  
; LENGTH: 2517  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-864-636A-96  
Alignment Scores:  
Pred. No.: 10.9 Length: 2517  
Score: 80.00 Matches: 55  
Percent Similarity: 42.13% Conservative: 20  
Best Local Similarity: 30.90% Mismatches: 58  
Query Match: 11.30% Indels: 46  
DB: 11 Gaps: 10  
US-10-087-573-2 (1-141) x US-09-864-636A-96 (1-2517)  
QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1141 CTGGACCTTCGAACACACACCCCGAGGGGTGGCG-----CGCGCTACGGGGGAG 1194  
QY 21 ---ThrPheAspValMetArgGluAlaLeu----- 30  
Db 1195 TGGACGGAGGACGCCGCCACCGGCGCTCTCTCGAGAGGCTCATCGGAACCTCCTT 1254  
QY 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1255 AAGCGCTCGAGGGGGAGGAGAGCTCTTGGCTCTACACGAGGTGGAAGAGCCCTC 1314  
QY 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 1315 TCCCGGTCTCTGGCCCATATGGA-GGCCACGCGGGGTGGCTTGGACGTGC----- 1364  
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArg----- 78  
Db 1365 CTATCTCAGGGCTTGTCCCT---GGAGGTGCCGAGGAGATCCCGCTCGAGCCGA 1421  
QY 79 -----ProGln-----SerThrLysSer 84  
Db 1422 GGTCTTCCGCTGGCGGCGCACCCCTTCAACTCCCGGACACAGCTGGAAGGCT 1481  
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGlu-MetAsnLysThrIleSerGlnG 104  
Db 1482 CTCTTTGACGAGCTAGGCTTCCCGCATCGGACGAGACGAGAGACCGCAAGCGCTC 1541  
QY 104 u---SerAlaArgValAsnHisArgLeuProGlyHisProLeuLeuGlyLysArgAl 123  
Db 1542 CACCAGCGCGCGCTCTCGAGGCGCTTCCGAGGCGCCACCCCATCTGTGAGAGATCCT 1601  
QY 123 agLutyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1602 GCAGTAC---CGGAGGCTCACCAAGCTGAAGAGACCTACGTGGACCCCTTG 1650  
RESULT 28  
US-09-758-282-141  
; Sequence 141, Application US/09758282  
; Publication No. US20030134349A1  
; GENERAL INFORMATION:  
; APPLICANT: Ma, Wu-Po  
; APPLICANT: Lyamichev, Victor I.  
; APPLICANT: Kaiser, Michael W.  
; APPLICANT: Lyamicheva, Natalie E.  
; APPLICANT: Allawi, Hatim T.  
; APPLICANT: Schaefer, James J.  
; APPLICANT: Nerri, Bruce P.  
; TITLE OF INVENTION: Improved Enzymes for the Detection of Specific Nucleic  
; FILE REFERENCE: FORS-04323  
; CURRENT APPLICATION NUMBER: US/09/758.282

APPLICANT: Kwiatkowski, Jr., Robert W.  
APPLICANT: Lukowiak, Andrew A.  
APPLICANT: Lyamichev, Victor  
APPLICANT: Lyamacheva, Natalie E.  
APPLICANT: Ma, WuPo  
APPLICANT: Neri, Bruce P.  
APPLICANT: Olson, Sarah M.  
APPLICANT: Olson-Munoz, Marilyn C.  
APPLICANT: Schaefer, James J.  
APPLICANT: Szrzipczynski, Zbigniew  
APPLICANT: Takova, Tssetska Y.  
APPLICANT: Thompson, Lisa C.  
APPLICANT: Thompson, Kevin L.  
APPLICANT: Vedvik, Kevin L.  
TITLE OF INVENTION: RNA Detection Assays  
FILE REFERENCE: FORS-06666  
CURRENT APPLICATION NUMBER: US/10/084,839  
CURRENT FILING DATE: 2002-02-26  
NUMBER OF SEQ ID NOS: 4004  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 96  
LENGTH: 2517  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic  
US-09-758-282-141

Alignment Scores:  
Pred. No.: 10.9 Length: 2517  
Score: 80.00 Matches: 55  
Percent Similarity: 42.13% Conservativeness: 20  
Best Local Similarity: 30.90% Mismatches: 58  
Query Match: 11.30% Indels: 46  
DB: 12 Gaps: 10

US-10-087-573-2 (1-141) x US-09-758-282-141 (1-2517)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1141 CTGGACCTTCGAACACACCCCGAGGGGTGGCG-----CGGCGCTACGGGGGGGAG 1194  
QY 21 ---ThrPheAspValMetArgGluAlaLeuLeu----- 30  
Db 1195 TGGACGGAGGAGCGCGCCACCGGGCCCTCTCTCGGAGAGGCTCCATCGGAACCTCCTT 1254  
QY 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1255 AAGCGCTCGAGGGGAGGAGAGCTCTTTGGCTCTACACGAGGTGGAAAGCCCTC 1314  
QY 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 1315 TCCCGGGTCTGGGCCCATATGGA-GGCCACGGGGGTGGCGTGCACGTGGC----- 1364  
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArg----- 78  
Db 1365 CTATCTCAGGGCTTGTCCT---GGAGGTGGCGGAGAGATCGCCGCTCGAGGCGGA 1421  
QY 79 -----ProGln-----SerThrLysSer 84  
Db 1422 GGTCTTCGCTGGCGCGCCACCCCTTCAACTCTCACTCCCGGGACCACTGGAAAGGT 1481  
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGlu-MetAsnLysThrIleSerGlnG 104  
Db 1482 CTCTTTGACGAGCTAGGGCTTCCGCCCATATGGA-GGCCACGGGGGTGGCGTGCACGTGC 1541  
QY 104 u---SerAlaArgValAsnHisArgLeuProGlyHisProLeuLeuGlyLysArgAl 123  
Db 1542 CACCAGCGCGCGTCTCGGAGGCGCTCCGCGAGGCGCCACCCCATCGTGGAGAGATCCT 1601  
QY 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1602 GCAGTAC---CGGAGGCTCACCAGCTGAAGAGCACCTACGTGGACCCCTTG 1650

RESULT 29

US-10-084-839-96  
Sequence 96, Application US/10084839  
Publication No. US20030186238A1  
GENERAL INFORMATION:  
APPLICANT: Third Wave Technologies  
APPLICANT: Allawi, Hatim  
APPLICANT: Argue, Brad T.  
APPLICANT: Bartholomay, Christian T.  
APPLICANT: Chehak, LuAnne  
APPLICANT: Curtis, Michelle L.  
APPLICANT: Eis, Peggy S.  
APPLICANT: Hall, Jeff G.  
APPLICANT: Ip, Hon S.  
APPLICANT: Ji, Lin  
APPLICANT: Kaiser, Michael

Alignment Scores:  
Pred. No.: 10.9 Length: 2517  
Score: 80.00 Matches: 55  
Percent Similarity: 42.13% Conservativeness: 20  
Best Local Similarity: 30.90% Mismatches: 58  
Query Match: 11.30% Indels: 46  
DB: 12 Gaps: 10

US-10-087-573-2 (1-141) x US-10-084-839-96 (1-2517)

QY 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1141 CTGGACCTTCGAACACACCCCGAGGGGTGGCG-----CGGCGCTACGGGGGGGAG 1194  
QY 21 ---ThrPheAspValMetArgGluAlaLeuLeu----- 30  
Db 1195 TGGACGGAGGAGCGCGCCACCGGGCCCTCTCTCGGAGAGGCTCCATCGGAACCTCCTT 1254  
QY 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1255 AAGCGCTCGAGGGGAGGAGAGCTCTTTGGCTCTACACGAGGTGGAAAGCCCTC 1314  
QY 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 1315 TCCCGGGTCTGGGCCCATATGGA-GGCCACGGGGGTGGCGTGCACGTGC----- 1364  
QY 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArg----- 78  
Db 1365 CTATCTCAGGGCTTGTCCT---GGAGGTGGCGGAGAGATCGCCGCTCGAGGCGGA 1421  
QY 79 -----ProGln-----SerThrLysSer 84  
Db 1422 GGTCTTCGCTGGCGCGCCACCCCTTCAACTCTCACTCCCGGGACCACTGGAAAGGT 1481  
QY 85 ProGluLeuArgGluLeuSerArgLysIleArgGlu-MetAsnLysThrIleSerGlnG 104  
Db 1482 CTCTTTGACGAGCTAGGGCTTCCGCCCATATGGA-GGCCACGGGGGTGGCGTGCACGTGC 1541  
QY 104 u---SerAlaArgValAsnHisArgLeuProGlyHisProLeuLeuGlyLysArgAl 123  
Db 1542 CACCAGCGCGCGTCTCGGAGGCGCTCCGCGAGGCGCCACCCCATCGTGGAGAGATCCT 1601  
QY 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1602 GCAGTAC---CGGAGGCTCACCAGCTGAAGAGCACCTACGTGGACCCCTTG 1650



RESULT 30  
US-09-547-267-8/c  
; Sequence 8, Application US/09547267  
; Patent No. US20020147371A1  
; GENERAL INFORMATION:  
; APPLICANT: Hohmann, Hans-Peter  
; APPLICANT: Pasamontes, Luis  
; APPLICANT: Tessier, Michel  
; APPLICANT: van Loon, Adolphus  
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/547,267  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/660,645  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pokras, Bruce A.  
; REGISTRATION NUMBER: 32,748  
; REFERENCE/DOCKET NUMBER: RAN 6002/170  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 235-5801  
; TELEFAX: (201) 235-2363  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1149 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-547-267-8  
Alignment Scores:  
Pred. No.: 4.45 Length: 1149  
Score: 79.50 Matches: 35  
Percent Similarity: 42.15% Conservative: 16  
Best Local Similarity: 28.93% Mismatches: 50  
Query Match: 11.23% Indels: 20  
DB: 10 Gaps: 4  
US-10-087-573-2 (1-141) x US-09-547-267-8 (1-1149)  
Qy 37 ArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArg-----51  
Db 590 CGGTATCTCGATCAGATCGGGTGGAGTCGAAGGGCAGCAGATAGATGAAGCGGTAC 531  
Qy 52 ---ValLeuProGlyThrGlyAlaSerAlaAlaAlaThrValThrProGlyGlyAla 70  
Db 530 CCGTCCATCTCGGAAACGGTCGCTCATGATCATCGGGCGCTCGACGCCATCGGGGGCG 471  
Qy 71 SerMetLysLeuLysProProArg-----ProLysSerThrLysSerProgl 86  
Db 470 TCGTCTCGATCTCGACGCCACCAATTCGGAACACCCACCGTCAGGTGCGGGGTCTCG 411  
Qy 86 uLeuArgGluLeuSerArgIysIleAlaGluMetAsnLysThrIleSerGlnGluSerAl 106  
Db 410 A---CGGCACCACCGGGCGTGCATCACCAGCAGCCCTCGATCCGCGAGCGCTCGTACG 354  
Qy 106 aArgValAsnHisArg-----LeuProGluGlyHisProLe 118

Db 353 GTCGCGCGGTATCGTCCAGGTCGCACATGCGTATTCCACCGCAGATCGACACCTGC 294  
Qy 118 uLeuGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValas 138  
Db 293 AGCAGCCCGATCAGCGCGCGCGCTCGATCGAGCCATAGCTGTCTGTCAGGCGCGGAA 234  
Qy 138 n 138  
Db 233 T 233  
RESULT 31  
US-10-156-761-2702  
; Sequence 2702, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: HORIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2702  
; LENGTH: 1668  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1668)  
US-10-156-761-2702  
Alignment Scores:  
Pred. No.: 7.28 Length: 1668  
Score: 79.50 Matches: 38  
Percent Similarity: 43.51% Conservative: 19  
Best Local Similarity: 29.01% Mismatches: 59  
Query Match: 11.23% Indels: 15  
DB: 14 Gaps: 7  
US-10-087-573-2 (1-141) x US-10-156-761-2702 (1-1668)  
Qy 14 AsnArgProThrPheGlyGluThrPheAspValMetArgGluAlaLeuArgValLys 33  
Db 456 CATCGGCCA-----GGAATTGGGGCCCGTGGGGCGTGGCGTCAACTCGCCCTGGGC 509  
Qy 34 SerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeu 53  
Db 510 GTCGCGCAGCGTTCAGCGCCTTCAGCGCTTCGCGCTGGCGGTCAA-----CATCG---CTT 560  
Qy 54 ProGlyThrGly-----AlaSerAlaIleAlaAlaThrValThrProLysGlyAla 70  
Db 561 CTGGGCCCGGGTTCGTCTCCGACCCGCGCCCTGTGTGAGGACACTCGCGAGGCGG 620  
Qy 71 SerMetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeu 90  
Db 621 CGGTCCCGCTCAAGTCCAGAGACATGGCGCCACACCCCGTGCCTCGACCCG 680  
Qy 91 SerArgLysIleArgGluMetAsnLysThrIleSerGlnGluSerAlaArgValAsnHis 110  
Db 681 GTTGGCGGT---CGCCGAGGAGCAGCATGTGTGAGGTGCGCCCTGCACAGCGAGCTGAA 737  
Qy 111 ArgLeuPro-----GluGlyHisProLeuLeuGluLysArgAlaGluTyrPheArgHis 128

[illegible]

[illegible]

SEQ ID NO 1558	42.13%	Conservative: 21
LENGTH: 1860	30.34%	Mismatches: 58
TYPE: DNA	11.16%	Indels: 46
ORGANISM: Streptomyces avermitilis	11	Gaps: 10
FEATURE:		
NAME/KEY: CDS		
LOCATION: (1)..(1860)		
US-10-156-761-1558		
Alignment Scores:		
Pred. No.:	9.69	Length: 1860
Score:	79.00	Matches: 39
Percent Similarity:	40.31%	Conservative: 13
Best Local Similarity:	30.23%	Mismatches: 41
Query Match:	11.16%	Indels: 36
DB:	14	Gaps: 8
US-10-087-573-2 (1-141) x US-10-156-761-1558 (1-1860)		
QY	4 ThrSerThrThrThrAsnPheValAla	-----GluAsn 14
DB	488 ACACAGACCACGACGCGTACGCGCGCGCGGTGAACCGCGAGACGCCCTCGACGAAC	429
QY	15 ArgProThrPheGly	-----GluThrPheAspValMetArgGluAlaLeuLeu 30
DB	428 CAGCCGATCTTCGACGTCCCGCGTGGAGACCGGACGATCTCGCGCGCGGACCGCG	369
QY	31 ArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHis	50
DB	368 CGGGTCAGCGGGGACCGACGCTGCGCGGTCTCGCG-----GCCAGCAGTCG	321
QY	51 ArgValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrValProLysGlyAla	70
DB	320 -----TCGACCGCGCGCGGTGATCCAGTTGGTGACGGCGCGCGGTGC	273
QY	71 SerMetLysLeuLysPro	-----ProArgProGlnSerThrLysSerProGluLeu 87
DB	272 AGCATGGAT--TCGCCACGCGCGGCGCGCGCGCGCACACGATCAGCCCGCGGTG	216
QY	88 ArgGluLeuSerArgLysIleArgGluMetAsnLysThrIle	-----101
DB	215 AGG-----GGAGTCGGGTGCCGCGCGGTCCACCGCTCGACGGCGGCGCGG	162
QY	102 -----SerGluSerAlaArg	107
DB	161 CAGAACGCGACGCGCGGAAACGCGCAGG	135
RESULT 37		
US-09-864-636A-431		
Sequence 431, Application US/09864636A		
Publication No. US20030104378A1		
GENERAL INFORMATION:		
APPLICANT: Third Wave Technologies		
APPLICANT: Allwai, Hatim		
APPLICANT: Bartholomay, Christian		
APPLICANT: Chehak, LuAnne		
TITLE OF INVENTION: Detection of RNA Sequences		
FILE REFERENCE: FORS-04944		
CURRENT APPLICATION NUMBER: US/09/864.636A		
CURRENT FILING DATE: 2002-10-15		
NUMBER OF SEQ ID NOS: 2640		
SOFTWARE: PatentIn version 3.0		
SEQ ID NO 431		
LENGTH: 2508		
TYPE: DNA		
ORGANISM: Artificial Sequence		
FEATURE:		
OTHER INFORMATION: Synthetic		
US-09-864-636A-431		
Alignment Scores:		
Pred. No.:	14.4	Length: 2508
Score:	79.00	Matches: 54

Alignment Scores:		
Pred. No.:	14.4	Length: 2508
Score:	79.00	Matches: 54

; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 73  
; LENGTH: 2517  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-864-636A-73

Alignment Scores:  
Pred. No.: 14.5 Length: 2517  
Score: 79.00 Matches: 54  
Percent Similarity: 42.13% Conservative: 21  
Best Local Similarity: 30.34% Mismatches: 58  
Query Match: 11.16% Indels: 46  
Gaps: 10

US-10-087-573-2 (1-141) x US-09-864-636A-73 (1-2517)

Qy 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1141 CTGGACCCCTTGAACACACCCCGAGGGGTGGCG-----CGGCGCTACGGGGGGGAG 1194  
Qy 21 ---ThrPheAspValMetArgGluAlaLeu----- 30  
Db 1195 TGGACGGAGGACCCCGCCACCGGGCCCTCTCTCGAGAGGCTCCATCGGAACCTCCTT 1254  
Qy 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1255 AAGCGCTCGAGGGGGAGGAGAGCTCTTTGGCTCTACCCAGAGGTGGAAGAGCCCTC 1314  
Qy 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60  
Db 1315 TCCCGGCTCTGGCCCATATGGA-GGCCACGGGGGTGGCGCTGGAGCTGGC----- 1364  
Qy 61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProArg----- 78  
Db 1365 CTATCTCAGGCTTTGCTCT---GGAGGTGGCGAGGAGATCGCCGCTCGAGGCCGA 1421  
Qy 79 -----ProGln-----SerThrLysSer 84  
Db 1422 GGTCTTCGCGCTGGCGGCCACCCCTTCAACTCCCGGAGCAGCTGGAAAGGGT 1481  
Qy 85 ProGluLeuArgGluLeuSerArgLysAlaLeuArgGlu-MetAsnLysThrIleSerGlnG 104  
Db 1482 CTTCTTTGACGAGTAGGGCTTCCGCCATCGGACGAGACGAGAGACCGGCAAGCGCTC 1541  
Qy 104 u---SerAlaArgValAsnHisArgLeuProGlyHisProLeuLeuGlyLysArgAl 123  
Db 1542 CACCAGCGCGCGCTCTCGAGGCGCTCCGAGGCGCCATCCGAGGAGGAGGAGGAGGAGG 1601  
Qy 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1602 CGAGTAC---CGGAGCTCACCAGCTGAAGAGCCTCATATTGACCCCTTG 1650

#### RESULT 40

US-09-864-636A-87  
; Sequence 87, Application US/09864636A  
; Publication No. US20030104378A1  
; GENERAL INFORMATION:  
; APPLICANT: Third Wave Technologies  
; APPLICANT: Allwai, Hatim  
; APPLICANT: Bartholomay, Christian  
; APPLICANT: Chehak, LuAnne  
; TITLE OF INVENTION: Detection of RNA Sequences  
; FILE REFERENCE: FORS-04944  
; CURRENT APPLICATION NUMBER: US/09/864,636A  
; NUMBER OF SEQ ID NOS: 2640  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 87  
; LENGTH: 2517  
; TYPE: DNA

; FILE REFERENCE: FORS-06666  
; CURRENT APPLICATION NUMBER: US/10/084,839  
; CURRENT FILING DATE: 2002-02-26  
; NUMBER OF SEQ ID NOS: 4004  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 431  
; LENGTH: 2508  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-10-084-839-431

Alignment Scores:  
Pred. No.: 14.4 Length: 2508  
Score: 79.00 Matches: 54  
Percent Similarity: 42.13% Conservative: 21  
Best Local Similarity: 30.34% Mismatches: 58  
Query Match: 11.16% Indels: 46  
Gaps: 10

US-10-087-573-2 (1-141) x US-10-084-839-431 (1-2508)

Qy 1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20  
Db 1150 CTGGACCCCTTGAACACACCCCGAGGGGTGGCG-----CGGCGCTACGGGGGGGAG 1203  
Qy 21 ---ThrPheAspValMetArgGluAlaLeu----- 30  
Db 1204 TGGACGGAGGACCCCGCCACCGGGCCCTCTCTCGAGAGGCTCCATCGGAACCTCCTT 1263  
Qy 31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41  
Db 1264 AAGCGCTCGAGGGGGAGGAGAGCTCTTTGGCTCTACCCAGAGGTGGAAGAGCCCTC 1323  
Qy 42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGly----- 57  
Db 1324 TCCCGGCTCTGGCCCATATGGA-GGCCACCGGGGTAGCGGGGACGTGGCTTCACTTCA 1382  
Qy 58 -----AlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeu 74  
Db 1383 GCGCCCTTCCCTGGAGCTTGGCGAGGAGATCGCGCTCGAGGAGGAGGTCTTCGCTT 1442  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeu-Arg---GluLeuSerArg-- 92  
Db 1443 GCGCGGCCACCCCTTCAA-----CCTCAACTCCCGGGACCCAGCTGGAAAGGGT 1490  
Qy 93 -----LysIleArgGluMetAsnLysThrIleSerGlnG 104  
Db 1491 GCTCTTTGACGAGCTTAGGCTTCCCAAGTTGAAGAAGACGAAGAAGACAGGCAAGCGCTC 1550  
Qy 104 u---SerAlaArgValAsnHisArgLeuProGlyHisProLeuLeuGlyLysArgAl 123  
Db 1551 CACCAGCGCGCGGTGCTGGAGGCGCTACGGAGGCGCCATCCGAGGAGGAGGAGGAGGAG 1610  
Qy 123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140  
Db 1611 CGAGTAC---CGGAGCTCACCAGCTCAAGACACCTACGTGGAGGAGGAGGAGGAGG 1659

#### RESULT 39

US-09-864-636A-73  
; Sequence 73, Application US/09864636A  
; Publication No. US20030104378A1  
; GENERAL INFORMATION:  
; APPLICANT: Third Wave Technologies  
; APPLICANT: Allwai, Hatim  
; APPLICANT: Bartholomay, Christian  
; APPLICANT: Chehak, LuAnne  
; TITLE OF INVENTION: Detection of RNA Sequences  
; FILE REFERENCE: FORS-04944  
; CURRENT APPLICATION NUMBER: US/09/864,636A  
; NUMBER OF SEQ ID NOS: 2640

```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-864-636A-87

```

```

Alignment Scores:
Pred. No.:      14.5      Length:      2517
Score:          79.00      Matches:      54
Percent Similarity: 42.13% Conservative: 21
Best Local Similarity: 30.34% Mismatches: 58
Query Match:      11.16% Indels: 46
DB:              11      Gaps: 10

```

US-10-087-573-2 (1-141) x US-09-864-636A-87 (1-2517)

```

QY      1 MetGluSerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGlu 20
DB      1141 CTGGACCTTCGAACACACCCCGGGGGTGGCG-----CGGCGTACGGGGGGGAG 1194
QY      21 ---ThrPheAspValMetArgGluAlaLeu----- 30
DB      1195 TGGACGGAGGACCGCCGCCACCGGGCCCTCTCTCGGAGAGGCTCCATCGGAACCTCTCTT 1254
QY      31 ---ArgValLysSerSerGluArgLeuAlaMetLeu----- 41
DB      1255 AAGCCCTCGAGGGGAGGAGAACTCTTTGGCTCTACACGAGGTGGAAAAGCCCTC 1314
QY      42 ---ArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAla 60
DB      1315 TCCCGGGTCTGGCCCATATGGA-GGCCACGGGGGTGGCGGGACGTGGC----- 1364
QY      61 IleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysProProArg----- 78
DB      1365 CTATCTCAGGGCTTGTCCCT---GGAGGTGGCCGAGGAGATCGCCCGCTCGAGGCCGA 1421
QY      79 -----ProGln-----SerThrLysSer 84
DB      1422 GGTCTTCGCGCTGGCGGCCACCCCTTCAACCTCAACTCCCGGGACAGCTGGAAGGCT 1481
QY      85 ProGluLeuArgGluLeuSerArgLysIleArgGlu-MetAsnLysThrIleSerGlnGln 104
DB      1482 CCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAACCGGCAAGCGCTC 1541
QY      104 u---SerAlaArgValAsnHisArgLeuProGlyHisProLeuLeuGluLysArgAl 123
DB      1542 CACCACGGCGCGCTCTCGGAGGCCCTCCGCGAGGCCACCCCTCATCGTGAGAGATCCT 1601
QY      123 aGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnArgLeu 140
DB      1602 GCAGTAC---CGGGAGCTCACCAGCTGAAGAGACCTACATTGACCCCTTG 1650

```

Search completed: November 17, 2003, 15:54:59  
Job time : 3118 secs

29: gb\_gss2:\*

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 17, 2003, 12:52:44 ; Search time 1928 Seconds  
(without alignments)  
1777.454 Million cell updates/sec

Title: US-10-087-573-2  
Perfect score: 708  
Sequence: 1 MESTSTTTNFVAENRPTFG.....RAEYFRLRLSLKSGQVRLI 141

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\*\_p2n.model DEV=xlh  
-Q=/cgn2\_1/USPRO\_spool/US10087573/runat\_14112003\_103610\_20685/app\_query.fasta\_1.327  
-DB=EST -QMT=fastap -SUFFIX=rat -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10087573 @CNG 1 1 2135 @runat\_14112003\_103610\_20685 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREAS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hcc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est6:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_lus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	101.5	14.3	665	14	CD468520	LeukoS3_3
C 2	96.5	13.6	1469	29	AG065308	Pan trogl
C 3	95.5	13.5	431	12	BI961232	MONO1_7_G
C 4	95.5	13.5	574	14	CD467780	LeukoS1_6
C 5	95.5	13.5	646	14	CD536579	LeukoN6_8
C 6	95.5	13.5	654	10	BE513677	601315412
C 7	95.5	13.5	663	14	CD468466	LeukoS3_3
C 8	95.5	13.5	671	14	CD467297	LeukoS1_3
C 9	95.5	13.5	700	14	CD468492	LeukoS3_3
C 10	95.5	13.5	729	14	CD535377	LeukoN5_2
C 11	95.5	13.5	738	14	CD528498	LeukoN3_2
C 12	95	13.4	881	12	BI248636	602993075
C 13	94.5	13.3	496	14	CD470400	LeukoS4_4
C 14	94	13.3	607	10	BI180114	602329694
C 15	94	13.3	669	12	BI181817	603033383
C 16	94	13.3	927	29	CC366660	PUH8B80TD
C 17	94	13.3	1172	12	BM909493	AGENCOURT
C 18	93.5	13.2	660	14	BY749167	BY749167
C 19	92.5	13.1	554	9	AV667883	AV667883
C 20	92.5	13.1	578	9	AV597776	AV597776
C 21	92.5	13.1	578	9	AV616040	AV616040
C 22	92.5	13.1	594	9	AV601755	AV601755
C 23	92.5	13.1	615	10	BE371432	RC0-FN014
C 24	92.5	13.1	641	10	BE574311	SNBST412
C 25	92.5	13.1	667	12	BP112124	BP112124
C 26	92.5	13.1	691	12	BI771910	603058924
C 27	92.5	13.1	875	12	BI757298	603029512
C 28	92.5	13.1	907	10	BF301271	602029713
C 29	92	13.0	1104	29	CNS0323B	Tetraodon
C 30	91.5	12.9	425	13	BY224833	BY224833
C 31	91.5	12.9	544	10	BE553346	BE553346
C 32	91.5	12.9	640	10	BE284887	BE284887
C 33	91.5	12.9	650	10	BE289370	BE289370
C 34	91.5	12.9	674	12	BI150185	602848591
C 35	91.5	12.9	727	10	BF581991	BF581991
C 36	91.5	12.9	768	12	BI696433	60345762
C 37	91.5	12.9	786	12	BI158106	60320741
C 38	91.5	12.9	835	12	BI839165	603086487
C 39	91.5	12.9	849	12	BI411683	BI411683
C 40	91.5	12.9	872	14	CB586950	CB586950
C 41	91.5	12.9	900	13	BQ952740	BQ952740
C 42	91.5	12.9	929	10	BF302274	BF302274
C 43	91.5	12.9	934	12	BI525023	602926430
C 44	91.5	12.9	961	12	BI904618	BI904618
C 45	91.5	12.9	968	12	BI108132	BI108132

ALIGNMENTS

RESULT 1  
LOCUS CD468520/c  
DEFINITION LeukoS3\_3\_D04.g1\_A025 Stimulated peripheral blood leukocytes S3  
ACCESSION CD468520  
VERSION CD468520.1  
KEYWORDS EST.  
SOURCE Equus caballus (horse)  
ORGANISM Equus caballus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
REFERENCE 1 (bases 1 to 665)

CD468520 665 bp mRNA linear EST 04-JUN-2003  
LeukoS3\_3\_D04.g1\_A025 Stimulated peripheral blood leukocytes S3  
Equus caballus cDNA clone LeukoS3\_3\_D04\_A025 5', mRNA sequence.  
CD468520  
CD468520.1 GI:31389788  
EST.  
Equus caballus (horse)  
Equus caballus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
REFERENCE 1 (bases 1 to 665)

Vandenplas, M., Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Moore, J.N., Liang, C., Sun, F., Sullivan, R., Shah, M. and Pratt, L.H.  
 An EST database from equine (*Equus caballus*) stimulated peripheral blood leukocytes  
 Unpublished  
 Other\_ESTs: LeukoS3\_3\_D04\_b1\_A025  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; tissue and RNA were prepared in the Department of Large Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: Sug5 (CTTCGCTCTAAAGTGGC).  
 Location/Qualifiers  
 1..665  
 /organism="Equus caballus"  
 /mol\_type="mRNA"  
 /strain="Tennessee walking horse"  
 /db\_xref="taxon:9796"  
 /clone="LeukoS3\_3\_D04\_A025"  
 /sex="female"  
 /tissue\_type="blood"  
 /cell\_type="leukocytes"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Stimulated peripheral blood leukocytes S3"  
 /note="Organ: circulatory system; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from equine peripheral blood leukocytes isolated from a healthy adult horse. The leukocytes were stimulated for 4 hr with 10 ng/ml E. coli O55:B5 LPS. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."  
 116 a 200 c 210 g 139 t  
 BASE COUNT  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3.79 Length: 665  
 Score: 101.50 Matches: 40  
 Percent Similarity: 44.4% Conservative: 16  
 Best Local Similarity: 31.7% Mismatches: 46  
 Query Match: 14.3% Indels: 24  
 DB: 14 Gaps: 5  
 US-10-087-573-2 (1-141) x CD468520 (1-665)  
 QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 DB 491 TCACGGGTGACGACAGCTTCCTCAGGGGCGACACCAGCCCGCTCAGGATCTCCAC 432  
 QY 23 AspValMetArgGluAlaLeuLeuArgVallySerSerGluArgLeuAlaMetLeuArg 42  
 DB 431 -----ACCATGACTGTGCAATCTCTGAGCCCACTGGCAATGACGTTA 390  
 QY 43 AlaLeuAlaGlyWetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 DB 389 TCATTG-----TGGGGCAGCAGCGATGTCACACAGGGGCTGTGTGCCACAG 339  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLeu 74  
 DB 338 ACCGTGGGGCGATTCTTGTCCACAGTCCAGTCTTGTCCCGGGGCGAGCACCAGGAAGGCC 279  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

278 CTCCCGCGTGGCTCACATACGAGGCGCCAGAACTTGGGGTTGACACACAG----- 225  
 QY 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProG1 114  
 DB 224 -----AAGCCACTGTCCAGGTGGTCTGGAGACGCGGCACATCTCTATACAC 177  
 QY 114 uGlyHisProLeuLeu 119  
 DB 176 TGGTCGCGCTTGGCTG 161  
 RESULT 2  
 AG065308 1469 bp DNA linear GSS 03-NOV-2001  
 LOCUS Pan troglodytes DNA, clone: PTB-054K10.R, genomic survey sequence.  
 ACCESSION AG065308  
 VERSION AG065308.1 GI:16617110  
 KEYWORDS GSS.  
 SOURCE Pan troglodytes (chimpanzee)  
 ORGANISM Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 REFERENCE 1  
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE BAC end sequences of Library PTB  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1469)  
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpbes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.  
 PRIMERS  
 Sequencing: M13Rev  
 LIBRARY Vector : pKS145  
 R.Site 1 : SacI  
 R.Site 2 : SacI.  
 FEATURES  
 source  
 1..1469  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="PTB-054K10.R"  
 /sex="male"  
 /cell\_type="lymphoblast"  
 /clone\_lib="PTB Chimpanzee Male BAC Library"  
 BASE COUNT 313 a 591 c 311 g 233 t 21 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 32.6 Length: 1469  
 Score: 96.50 Matches: 36  
 Percent Similarity: 40.44% Conservative: 19  
 Best Local Similarity: 26.47% Mismatches: 46  
 Query Match: 13.63% Indels: 35  
 DB: 29 Gaps: 6  
 US-10-087-573-2 (1-141) x AG065308 (1-1469)  
 QY 12 AlaGluAsnArgProThr-----PheGlyGluThrPheAspVal 24  
 DB 713 GCTCAGACCGCCCGAGCTACCTCAATCGCTACGACGCCACGACACTGC 772  
 QY 25 MetArgGluAlaLeuArgVallySerSerGluArgLeuAlaMetLeuArgAlaLeu 44



Db 773 CCACGCTACACATCTCAAAAGTATGATGACGAGCAGCTTCTGTCATGACTCCGAAGTGTC 832

Qy 45 AlaGlyMetCysGlyHisArgValLeuProGly 56

Db 833 GCG-----CGTCTCACTCCCTCGATTGAGTCGCCCTTCGATCTCCACGCC 877

Qy 57 GlyAlaSerAlaIleAlaAlaThrValThrProLysGlyAlaSerMetLysLeuLysPro 76

Db 878 GCGCGCGTGTCTCGCGGCTGCATCCGCCGCTACGTCAGTCCGCGC---AGCACCC 934

Qy 77 ProArgProGlnSerThrLysSerProGluLeu-----ArgGluLeuSerArgLys 93

Db 935 CCGCGCCCTCTCTCCGCCAGTCGTCCTCGATCCACCGCATCTCGTCGCCCTTCACGACGN 994

Qy 94 IleArgGluMet-----AsnLysThrIleSer 102

Db 995 GTACACCATCTTACCATCACACTCCGTCCTCGATACGACCTCCACACACGCGCATCC 1054

Qy 103 GlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeu 118

Db 1055 TCGCAATCTTACGCGCTAACCTGCTGCATCCACCGCGCGCCGACCACTA 1102

RESULT 3

LOCUS BI961232/c 431 bp mRNA linear EST 22-OCT-2001

DEFINITION MON01.7.G02.bi.A005 Monocytes (MON01) Equus caballus cDNA, mRNA

ACCESSION BI961232

VERSION BI961232.1 GI:16319435

KEYWORDS EST.

SOURCE Equus caballus (horse)

ORGANISM Equus caballus

REFERENCE 1 (bases 1 to 431)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

TITLE Gingle,A.R., Pratt,L.H. and Moore,J.N.

JOURNAL An EST database from equine (Equus caballus) monocytes

COMMENT Unpublished

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyMix or T7 sequencing primer, are presented as the reverse complement.

Seq primer: JEN REV

High quality sequence stop: 425

POLYA=No.

FEATURES

source

1..431

/organism="Equus caballus"

/mol\_type="mRNA"

/db\_xref="taxon:9796"

/cell\_type="isolated peripheral blood monocytes stimulated with E. coli lipopolysaccharide"

/clone\_lib="Monocytes (MON01)"

/note="Vector: pBluescript SK(-) from Lambda ZapII; Site 1: XhoI; Site 2: EcoRI. The library was made from polyA RNA in the cloning vector lambda ZapII. Clones to be sequenced were prepared by mass excision."

BASE COUNT 78 a 127 c 147 g 79 t

ORIGIN

Alignment Scores:

Pred. No.: 8.52 Length: 431

Score: 95.50 Matches: 39

Percent Similarity: 43.65% Conservative: 16

Best Local Similarity: 30.95% Mismatches: 47

Query Match: 13.49% Indels: 24

DB: 12 Gaps: 5

US-10-087-573-2 (1-141) x BI961232 (1-431)

Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22

Db 349 TCCAGGCTGACGACACTTCCCTCAGGCGCAGCACCAGGCCCCCGTCAGGATCTCCAC 290

Qy 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42

Db 289 -----ACCATGACTGTGCAATCTCCGAGCCACTGGCAATGACGTTA 248

Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 247 TCATTG-----TCGCGGCACAGCGCATGTCAGCAGCAGGGCTGTGTGGCCACAG 197

Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74

Db 196 ACGTGGCGCATCTTGTCCACACGTCACGTCCTGCCAGGCGCAGCACCAGGAAGGCC 137

Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94

Db 136 CTTCCCGCGTGGCTCCACATATGAGGCGCACAACCTGGGGTTGACAGCAG----- 83

Qy 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114

Db 82 -----AAGCACGTCTCCAGGTGGTCTGCGAGACGCGCACATCTCTCATAACAC 35

Qy 114 uGlyHisProLeuLeu 119

Db 34 TGGTCGCGCTTGGCTG 19

RESULT 4

LOCUS CD467780/c 574 bp mRNA linear EST 04-JUN-2003

DEFINITION LeukoS1\_6\_E02.g1\_A023 Stimulated peripheral blood leukocytes S1

ACCESSION CD467780

VERSION CD467780.1 GI:31389048

KEYWORDS EST.

SOURCE Equus caballus (horse)

ORGANISM Equus caballus

REFERENCE 1 (bases 1 to 574)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

TITLE Vandenplas,M., Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Moore

JOURNAL An EST database from equine (Equus caballus) stimulated peripheral blood leukocytes

COMMENT Unpublished

Other ESTs: LeukoS1\_6\_E02.bl\_A023

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; tissue and RNA were prepared in the Department of Large Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: SugS (CTTCTGCTCTAAAGCTCGC).

FEATURES

source

1..574

/organism="Equus caballus"

/mol\_type="mRNA"

/strain="Belgian draft"

/db\_xref="taxon:9796"  
 /clone="LeukoS1\_6\_E02\_A023"  
 /sex="female"  
 /tissue\_type="leukocytes"  
 /cell\_type="leukocytes"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Stimulated peripheral blood leukocytes S1"  
 /note="Organ: circulatory system; Vector: pME18S-FL3;  
 Site 1: XhoI; Site 2: XhoI; The library was prepared from  
 polyA+ RNA from equine peripheral blood leukocytes  
 isolated from a healthy adult horse. The leukocytes were  
 stimulated for 4 hr with 10 ng/ml E. coli O55:B5 LPS.  
 Double-stranded cDNA was cloned unidirectionally into  
 different DraIII sites of the pME18S-FL3 vector (5-prime  
 DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG  
 ). XhoI excises the cDNA insert."

BASE COUNT 99 a 179 c 176 g 120 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 12.3 Length: 574  
 Score: 95.50 Matches: 39  
 Percent Similarity: 43.65% Conservative: 16  
 Best Local Similarity: 30.95% Mismatches: 47  
 Query Match: 13.49% Indels: 24  
 DB: 14 Gaps: 5

US-10-087-573-2 (1-141) x CD467780 (1-574)

QY 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 486 TCCAGGGTGACGACAGCTTCCCTCAGGGGACGACCCAGGGCCCGCTCAGGGATCTCCAC 427  
 QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 Db 426 -----ACCATGACTGTGCATCTCTCGAGCCACTGCGCATGACGTTA 385  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 Db 384 TCATTG-----TGGGGCCACAGGGGATCTCCAGCACAGGGGCTGTGTGGCCACAG 334  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 Db 333 ACCGTGGGGCGATCTGTGCACAGTCAGTCTTCCAGGGGACGACCGAGGAGGCC 274  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 Db 273 CTTCCCGCGCTGCCTCACATATGAGGGCCAGCACTTGGGGTTGACAGCACAG----- 220  
 QY 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114  
 Db 219 -----AAGCCACTGTCCAGTGTCTGCGAGCGCCACATCTCTCATTAACAC 172  
 QY 114 uGlyHisProLeuLeu 119  
 Db 171 TGTGGCGCTTGGCTG 156

RESULT 5  
 CD536579/c  
 LOCUS  
 DEFINITION  
 Leukone\_8\_D06\_g1\_A028 Unstimulated peripheral blood leukocytes N6  
 Equus caballus cDNA clone Leukone\_8\_D06\_A028 5', mRNA sequence.  
 CD536579  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Equus caballus (horse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 1 (bases 1 to 646)  
 Vandenplas, M., Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Moore  
 J.N., Liang, C., Sun, F., Sullivan, R., Shah, M., and Pratt, L.H.  
 'A NEST database from equine (Equus caballus) unstimulated

peripheral blood leukocytes  
 Unpublished  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
 the Human Genome Center, University of Tokyo Institute of Medical  
 Science; tissue and RNA were prepared in the Department of Large  
 Animal Medicine, University of Georgia; sequencing done in the  
 Laboratory for Genomics and Bioinformatics, University of Georgia.  
 Sequence ends have been trimmed to exclude vector and regions below  
 phred quality 16. Three-prime sequences are presented as their  
 reverse complement and have been trimmed to exclude polyA.  
 Seq primer: Sug5 (CTTGTGCTTAAAGACTGCG).  
 Location/Qualifiers

FEATURES  
 source  
 1..646  
 /organism="Equus caballus"  
 /mol\_type="mRNA"  
 /strain="quaterhorse"  
 /db\_xref="taxon:9796"  
 /clone="Leukone\_8\_D06\_A028"  
 /sex="male"  
 /tissue\_type="blood"  
 /cell\_type="leukocytes"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Unstimulated peripheral blood leukocytes N6"  
 /note="Organ: circulatory system; Vector: pME18S-FL3;  
 Site 1: XhoI; Site 2: XhoI; The library was prepared from  
 polyA+ RNA from unstimulated equine peripheral blood  
 leukocytes isolated from a healthy adult horse.  
 Double-stranded cDNA was cloned unidirectionally into  
 different DraIII sites of the pME18S-FL3 vector (5-prime  
 DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG  
 ). XhoI excises the cDNA insert."

BASE COUNT 113 a 195 c 202 g 136 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 14.3 Length: 646  
 Score: 95.50 Matches: 39  
 Percent Similarity: 43.65% Conservative: 16  
 Best Local Similarity: 30.95% Mismatches: 47  
 Query Match: 13.49% Indels: 24  
 DB: 14 Gaps: 5

US-10-087-573-2 (1-141) x CD536579 (1-646)

QY 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 491 TCCAGGGTGACGACAGCTTCCCTCAGGGGACGACCCAGGGCCCGCTCAGGGATCTCCAC 432  
 QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 Db 431 -----ACCATGACTGTGCATCTCTCGAGCCACTGCGCATGACGTTA 390  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 Db 389 TCATTG-----TGGGGCCACAGGGGATCTCCAGCACAGGGGCTGTGTGGCCACAG 339  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 Db 338 ACCGTGGGGCGATCTTGTTCACAGCTCCAGTCTTGGCCAGGGGACGACCCAGGAGGCC 279  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 Db 278 CCTCCCGCGCTGCCTCACATATGAGGGCCAGCAACTTGGGGTTGACAGCACAG----- 225  
 QY 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114

Db 224 -----ANGCAGCTGTCCGAGTGGTGTCGGAGACGGGCACATCCTCAATAACAC 177

QY 114 uGlyHisProLeuEun 119  
|||||  
Db 176 TGGTCGGCGCTTGGCTG 161

RESULT 6  
BES13677/c  
LOCUS BES13677.1  
DEFINITION Homo sapiens CDNA clone IMAGE:3634202 5', mRNA linear EST 07-AUG-2000  
ACCESSION BE513677  
VERSION BE513677.1 GI:9720889  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 654)  
NIH-MGC http://mgc.nci.nih.gov/.  
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC) Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bs-research.nhlbi.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCW328 row: G column: O3  
High quality sequence stop: 654.

FEATURES  
Location/Qualifiers  
1..654  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3634202"  
/tissue\_type="Burkitt lymphoma"  
/lab\_host="DHIOB (phage-resistant)"  
/clone\_lib="NIH\_MGC\_8"  
/note="Organ: lymph; Vector: pORF7; Site:1: XhoI; Site:2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 114 A 201 C 207 G 132 U

ORIGIN

Alignment Scores:  
Pred. No.: 14.5 Length: 654  
Score: 95.50 Matches: 46  
Percent Similarity: 40.13% Conservative: 17  
Best Local Similarity: 29.30% Mismatches: 50  
Query Match: 13.49% Indels: 44  
DB: 10 Gaps: 7

US-10-087-573-2 (1-141) x BE513677 (1-654)

QY 8 ThrAsnPhelAlaGluaSnArgProThrPheGlyGluThrPheaspValMetArg--- 26  
:::  
Db 476 ACACGCTGGTGTGGCCCTCCAGGCGTACGACGGGCTCCCGCAGGGCGCAGCATCAGGCC 417  
:::  
QY 27 -----GluAlaLeuArgVallySerSerGluArgLeuAlaMetLeuArg 42  
:::  
Db 416 CCATCTGGATCTCCCACCACCATGACTGTGCGAGTCCTCGGAGCCACTGGCAATGACGTGG 357  
:::

QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
:::  
Db 356 TCATTG-----TGGCGCACACGCGATCTTAGCACAGGGCTGTGTGGCCACAG 308  
:::

QY	59	-----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
Db	305	ACCGTGGGCGGCAATCTTGTCCACACGTCACAGTCTTGCCCGGCGACGACGAGGAAGGCC 246
QY	75	LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db	245	CCTGCCCGCG-----CTGGCCTCACAGATC 222
QY	95	ArgGluMetAsn-----LysThrIleSerGlnGlu-SerAlaArgVa 108
Db	221	AGGGCCACAACTTAGGTTGACAGCACAGAAAGCCACTGTCCCGAGGTGCTCTGTGAGACG 162
QY	108	lasnHisArgLeuProGluGlyHisProLeu-----Le 119
Db	161	CGCACATCTTCATAGCACTGGTCGGCCATGGCCGGCTCTCCAAACACGTGGCGGAACCT 102
QY	119	uGluLysArgAlaGluTyPheArgHisLeuArgSerLeuLysSerGln 135
Db	101	GCTGGAGCGGACCACTCGCGCGCTCATCTGTGTCGAGAGAGGACAG 53
RESULT 7		
LOCUS	CD468466/c	
DEFINITION	LeukoS3_3_G11_g1_A025 Stimulated peripheral blood leukocytes S3	
ACCESSION	Equus caballus cDNA clone LeukoS3_3_G11_A025 5', mRNA sequence.	
VERSION	CD468466	
KEYWORDS	CD468466.1 GI:31389734	
SOURCE	EST.	
ORGANISM	Equus caballus (horse)	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.	
AUTHORS	1 (bases 1 to 663) Vandenplas,M., Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Moore,J.N., Liang,C., Sun,P., Sullivan,R., Shah,M. and Pratt,L.H.	
TITLE	An EST database from equine (Equus caballus) stimulated peripheral blood leukocytes	
JOURNAL	Unpublished	
COMMENT	Other_ESTs: LeukoS3_3_G11.bt_A025 Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; tissue and RNA were prepared in the Department of Large Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA. Seq primer: Sug5 (CTCTGCTCTAAAAGCTCG).	
FEATURES	Location/Qualifiers	
source	1..563 /organism="Equus caballus" /mol_type="mRNA" /strain="Tennessee walking horse" /db_xref="taxon:9796" /clone="LeukoS3_3_G11_A025" /sex="female" /tissue_type="blood" /cell_type="leukocytes" /lab_host="DH10B-T1 phage-resistant E. coli" /clone_lib="Stimulated peripheral blood leukocytes S3" /note="Organ: circulatory system; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from equine peripheral blood leukocytes isolated from a healthy adult horse. The leukocytes were stimulated for 4 hr with 10 ng/ml P. coli 055:85 LPS	

Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5'-prime DraIII site is CACTGTGTG, 3'-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

BASE COUNT 118 a 198 c 205 g 142 t

ORIGIN

Alignment Scores:

Pred. No.:	14.8	Length:	663
Score:	95.50	Matches:	39
Percent Similarity:	43.65%	Conservative:	16
Best Local Similarity:	30.95%	Mismatches:	47
Query Match:	13.49%	Indels:	24
DB:	14	Gaps:	5

US-10-087-573-2 (1-141) x CD468466 (1-663)

Qy 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 501 TCAGGGTGACGACAGCTTCCTCAGGGGAGCAGCAGGCCCCCGTCAGGATCTCCAC 442

Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 Db 441 -----ACCATGACTGTGCAATCTCGAGGCACTGGCAATGACGTTA 400

Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 Db 399 TCATTG-----TGGGGGACAGGCGATCTCCAGCAGGGGCTGTGGCCACAG 349

Qy 59 -----SerAlaIleAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 Db 348 ACCGTGGCGCATTTCTTGCACAGCTTCAGTCTCCAGGGGAGCAGCAGGAGGCG 289

Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgLysSerArgLysIle 94  
 Db 288 CTCTCCCGCGTGGCTCACATATAGGGCCACGAACTTGGGGTTGACAGCAG----- 235

Qy 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114  
 Db 234 -----AAGCCACTGTCCAGGTGTCTCGAGAGCGGCACATCTCTATAACAC 187

Qy 114 uGlyHisProLeuLeu 119  
 Db 186 TGGTCGGCCTTGGCTG 171

RESULT 8  
 CD467297/c  
 LOCUS  
 DEFINITION Leukos1\_3\_A07.g1\_A023 Stimulated peripheral blood leukocytes S1  
 Equus caballus cDNA clone Leukos1\_3\_A07\_A023 5', mRNA sequence.  
 ACCESSION CD467297  
 KEYWORDS CD467297.1 GI:31388565  
 SOURCE EST.  
 ORGANISM Equus caballus (horse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 REFERENCE 1 (bases 1 to 671)  
 AUTHORS Vandenplas, M., Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Moore, J.N., Liang, C., Sun, F., Sullivan, R., Shah, M. and Pratt, L.H.  
 TITLE An EST database from equine (Equus caballus) stimulated peripheral blood leukocytes  
 JOURNAL Unpublished  
 COMMENT Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; tissue and RNA were prepared in the Department of Large

Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sug5 (CTTCTGCTCTAAAGCTGGG).

FEATURES  
 source  
 location/Qualifiers  
 1..671  
 /organism="Equus caballus"  
 /mol\_type="mRNA"  
 /strain="Belgian draft"  
 /db\_xref="taxon:9796"  
 /clone="Leukos1\_3\_A07\_A023"  
 /sex="female"  
 /tissue\_type="blood"  
 /cell\_type="leukocytes"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Stimulated peripheral blood leukocytes S1"  
 /note="Organ: circulated peripheral blood leukocytes; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI. The library was prepared from polyA+ RNA from equine peripheral blood leukocytes isolated from a healthy adult horse. The leukocytes were stimulated for 4 hr with 10 ng/ml E. coli O55:B5 LPS. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5'-prime DraIII site is CACTGTGTG, 3'-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

BASE COUNT 119 a 200 c 211 g 141 t

ORIGIN

Alignment Scores:

Pred. No.:	15	Length:	671
Score:	95.50	Matches:	39
Percent Similarity:	43.65%	Conservative:	16
Best Local Similarity:	30.95%	Mismatches:	47
Query Match:	13.49%	Indels:	24
DB:	14	Gaps:	5

US-10-087-573-2 (1-141) x CD467297 (1-671)

Qy 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 Db 497 TCAGGGTGACGACAGCTTCCTCAGGGGAGCAGCAGGCCCCCGTCAGGATCTCCAC 438

Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 Db 437 -----ACCATGACTGTGCAATCTCGAGGCACTGGCAATGACGTTA 396

Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 Db 395 TCATTG-----TGGGGGACAGGCGATCTCCAGCAGGCGCTGTGTGGCCACAG 345

Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 Db 344 ACCGTGGGCGCATTTCTTGCACAGCTTCAGTCTTCCAGGGGAGCAGCAGGAGGCG 285

Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 Db 284 CTCTCCCGCGTGGCTCACATATAGGGCCACGAACTTGGGGTTGACAGCAG----- 231

Qy 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114  
 Db 230 -----AAGCCACTGTCCAGGTGTCTCGAGAGCGGCACATCTCTATAACAC 183

Qy 114 uGlyHisProLeuLeu 119  
 Db 182 TGGTCGGCCTTGGCTG 167

RESULT 9  
 CD468492/c  
 LOCUS  
 DEFINITION Leukos3\_3\_D08.g1\_A025 Stimulated peripheral blood leukocytes S3  
 Equus caballus cDNA clone Leukos3\_3\_D08\_A025 5', mRNA sequence.

ACCESSION CD468492  
 VERSION CD468492.1 GI:31389760  
 KEYWORDS EST.  
 SOURCE Equus caballus (horse)  
 ORGANISM Equus caballus  
 REFERENCE 1 (bases 1 to 700)  
 AUTHORS Vandenplas, M., Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Moore, J.N., Liang, C., Sun, F., Sullivan, R., Shah, M. and Pratt, L.H.  
 TITLE An EST database from equine (Equus caballus) stimulated peripheral blood leukocytes  
 JOURNAL Unpublished  
 COMMENT Other ESTs: LeukoS3\_3\_D08.b1.A025  
 Contact: Cordonnier-Pratt, M.-M.  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; tissue and RNA were prepared in the Department of Large Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: Sug5 (CTCTGCTCTAAAGCTGGC).  
 Location/Qualifiers  
 1..700  
 /organism="Equus caballus"  
 /mol\_type="mRNA"  
 /strain="Tennessee walking horse"  
 /db\_xref="taxon:9796"  
 /clone="LeukoS3\_3\_D08\_A025"  
 /sex="female"  
 /tissue\_type="blood"  
 /cell\_type="leukocytes"  
 /lab\_host="DH10B-Ti phage-resistant E. coli"  
 /clone\_lib="Stimulated peripheral blood leukocytes S3"  
 /note="Organ: circulatory system; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from equine peripheral blood leukocytes isolated from a healthy adult horse. The leukocytes were stimulated for 4 hr with 10 ng/ml E. coli O55:B5 LPS. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."  
 BASE COUNT 122 a 213 c 217 g 148 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 15.9 Length: 700  
 Score: 95.50 Matches: 39  
 Percent Similarity: 43.65% Conservative: 16  
 Best Local Similarity: 30.95% Mismatches: 47  
 Query Match: 13.43% Indels: 24  
 DB: 14 Gaps: 5  
 US-10-087-573-2 (1-141) x CD468492 (1-700)  
 QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGluThrPhe 22  
 Db TCCAGGTTGACGACAGCTTCCTCAGGGGCGAGCACCAGGCCCGCTAGGATCTCCAC 441  
 QY 23 AspValMetArgGluAlaLeuLeuArgValIysSerSerGluArgLeuAlaMetLeuArg 42  
 Db ACCATGACTGTGCAATCTCTCGAGCCCACTGGCAATGACGTTA 399  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58

Db 398 TCATTG-----TCGGGGCACCAGGGCGATCCAGCAGACGGGGCTGTGTGCCACAG 348  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 Db 347 ACCGTGGCGGCATTCTTGTCCACACGTCAGTCTTCCAGGAGGGGAGCAGCAGAGGCC 288  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 Db 287 CTTCCCGCGCTGGCCCTCACATATGAGGGGCCACGAATTTGGGGTTGACAGCAGCAG- 234  
 QY 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProG1 114  
 Db 233 -----AAGCCACTGTCCAGGTGCTTGGAGAGCGGCACATCTCTATAACAC 186  
 QY 114 uGlyHisProLeuLeu 119  
 Db 185 TGGTCGGCCTTGGCTG 170  
 RESULT 10  
 LOCUS CD535377/c  
 DEFINITION LeukONS\_2\_B09.g1\_A027 Unstimulated peripheral blood leukocytes N5  
 Equus caballus cDNA clone LeukONS\_2\_B09\_A027 5', mRNA sequence.  
 CD535377  
 ACCESSION CD535377.1 GI:31577792  
 VERSION EST.  
 KEYWORDS Equus caballus (horse)  
 SOURCE Equus caballus  
 ORGANISM Equus caballus  
 REFERENCE 1 (bases 1 to 729)  
 AUTHORS Vandenplas, M., Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Moore, J.N., Liang, C., Sun, F., Sullivan, R., Shah, M. and Pratt, L.H.  
 TITLE An EST database from equine (Equus caballus) unstimulated peripheral blood leukocytes  
 JOURNAL Unpublished  
 COMMENT Contact: Cordonnier-Pratt, M.-M.  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; tissue and RNA were prepared in the Department of Large Animal Medicine, University of Georgia; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: Sug5 (CTTCTGCTCTAAAGCTGGC).  
 Location/Qualifiers  
 1..729  
 /organism="Equus caballus"  
 /mol\_type="mRNA"  
 /strain="Dartmoor Pony"  
 /db\_xref="taxon:9796"  
 /clone="LeukONS\_2\_B09\_A027"  
 /sex="male"  
 /tissue\_type="blood"  
 /cell\_type="leukocytes"  
 /lab\_host="DH10B-Ti phage-resistant E. coli"  
 /clone\_lib="Unstimulated peripheral blood leukocytes N5"  
 /note="Organ: circulatory system; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from unstimulated equine peripheral blood leukocytes isolated from a healthy adult horse. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

BASE COUNT 129 a 225 c 225 g 150 t

ORIGIN

Alignment Scores:

Pred. No.: 16.7 Length: 729  
Score: 95.50 Matches: 39  
Percent Similarity: 43.65% Conservatve: 16  
Best Local Similarity: 30.95% Mismatches: 47  
Query Match: 13.49% Indels: 24  
DB: 14 Gaps: 5

US-10-087-573-2 (1-141) x CD535377 (1-729)

Qy 3 SerThrSerThrThrThrAsnPhValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 491 TCCAGGGTGACGACGCTTCCCTCAGGGCCAGCAGCCGCCCTCAGGATCTCCAC 432  
Qy 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 431 -----ACCATGACTGTGCAATCTCGAGCCACTGGCATGAGCTTA 390  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Db 389 TCATTG-----TGCAGGACACGAGCGATGTCACGACAGGGGCTGTGTGGCCACAG 339  
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 338 ACCGTGGGGGCACTTCTTGCACAGCTCCAGTCTTCCAGGGGAGGAGGCC 279  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 278 CTTCCCGCCGCTCGCTCACAATATGAGGCGCAGAACTTGGGGTTGACAGCAG----- 225  
Qy 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114  
Db 224 -----AAGCATCTGCCAGGTGGTCTCGAGACGCGCATCTCATAACAC 177  
Qy 114 uGlyHisProLeuLeu 119  
Db 176 TGGTGGCCTTGGCTG 161

RESULT 11

CD528498/c

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Comment

CD528498 738 bp mRNA linear EST 10-JUN-2003  
LeukN3\_2\_C10\_g1\_A025 Unstimulated peripheral blood leukocytes N3  
Equus caballus cDNA clone LeukN3\_2\_C10\_A025 5', mRNA sequence.  
CD528498  
CD528498.1 GI:31567120  
EST.  
Equus caballus (horse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
J.N., Liang, C., Sun, F., Sullivan, R., Shah, M. and Pratt, L.H.  
An EST database from equine (Equus caballus) unstimulated  
peripheral blood leukocytes  
Unpublished  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; tissue and RNA were prepared in the Department of Large  
Animal Medicine, University of Georgia; sequencing done in the  
Laboratory for Genomics and Bioinformatics, University of Georgia.  
Sequence ends have been trimmed to exclude vector and regions below  
Phred quality 16. Three-prime sequences are presented as their

reverse complement and have been trimmed to exclude polyA.  
Seq primer: Sug5 (CTTCTGCTCTAAAGCTGCG).

FEATURES  
source

1..738  
Location/Qualifiers  
/organism="Equus caballus"  
/mol\_type="mRNA"  
/strain="Tennessee walking horse"  
/db\_xref="taxon:9796"  
/clone="LeukN3\_2\_C10\_A025"  
/sex="female"  
/tissue\_type="blood"  
/cell\_type="leukocytes"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="Unstimulated peripheral blood leukocytes N3"  
/notes="Organ: circulatory system; Vector: pME18S-FL3;  
Site 1: XhoI; Site 2: XhoI; The library was prepared from  
polyA+ RNA from unstimulated equine peripheral blood  
leukocytes isolated from a healthy adult horse.  
Double-stranded cDNA was cloned unidirectionally into  
different DraIII sites of the pME18S-FL3 vector (5-prime  
DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG  
). XhoI excises the cDNA insert."  
BASE COUNT 131 a 225 c 229 g 153 t  
ORIGIN

Alignment Scores:  
Pred. No.: 17 Length: 738  
Score: 95.50 Matches: 39  
Percent Similarity: 43.65% Conservatve: 16  
Best Local Similarity: 30.95% Mismatches: 47  
Query Match: 13.49% Indels: 24  
DB: 14 Gaps: 5

US-10-087-573-2 (1-141) x CD528498 (1-738)

Qy 3 SerThrSerThrThrThrAsnPhValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 490 TCCAGGGTGACGACGCTTCCCTCAGGGCCAGCAGCCGCCCTCAGGATCTCCAC 431  
Qy 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
Db 430 -----ACCATGACTGTGCAATCTCGAGGACCATCTCATAACAC 389  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Db 388 TCATTG-----TGCAGGACACGAGCGATGTCACGACAGGGGCTGTGTGGCCACAG 338  
Qy 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 337 ACCGTGGGGGCACTTCTTGCACAGCTCCAGTCTTCCAGGGGAGGAGGCC 278  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 277 CTTCCCGCCGCTCGCTCACAATATGAGGCGCAGAACTTGGGGTTGACAGCAG----- 224  
Qy 95 ArgGluMetAsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGln 114  
Db 223 -----AAGCATCTGCCAGGTGGTCTCGAGACGCGCATCTCATAACAC 176  
Qy 114 uGlyHisProLeuLeu 119  
Db 175 TGGTGGCCTTGGCTG 160

RESULT 12

B1248636/c

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

B1248636 881 bp mRNA linear EST 17-JUL-2001  
602993075F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:5148690 5',  
mRNA sequence.  
B1248636  
B1248636.1 GI:14795221  
EST.  
Mus musculus (house mouse)  
Mus musculus

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 881)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov
COMMENT	Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LLAM11367 row: f column: 19 High quality sequence stop: 877. Location/Qualifiers 1. .881 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:5148690" /tissue_type="tumor, gross tissue" /dev_stage="7 months" /lab_host="DH10B" /lab_lib="NCI_COAP Mam5" /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"
BASE COUNT	168 a 251 c 279 g 183 t
ORIGIN	
Alignment Scores:	

```

Best Local Similarity: 27.74% Mismatches: 36
Query Match: 13.42% Indels: 60
DB: 12 Gaps: 9

US-10-087-573-2 (1-141) x B1248636 (1-881)

QY 3 SerThrSerThrThrThrAspPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 405 TCCAAAGGTGACGACAGGCTCCCGAAGGGCAGCACACGAGGCCCCATCCGGGATCTCCAC 346
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 23 AspValMetArgGluAlaLeuLeuArgValIysSerSerGluArgLeuAlaMetLeuArg 42
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 ACCATA-----ACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTG 304

QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 TCATTG-----TGTGGACACAGCGATGTCTAGCAGAGGGCAGTGTGGCGCGAG 253

QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 ACCAGGGCGCAGTTCTTGTCTACTCGTCCAGTCTTGTCTAGGGGTAGCACAGGAAGGCC 193

QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 CCTCCCCCA-----CTGGCCCTCACAGATC 169

QY 95 ArgGluMetAsn-----LysThrIleSerGln----- 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168 AGAGCGCATGAATCGGGGTTCACAGCGCAGAAGCCACTGTCCCAAGTGGTGTGTGAGACG 109

QY 104 -----GluSer 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 CGCACATCTCATAGCACTGGTCAGCCTTGGCTGGCTGTGCCAAACACGTGGCGGAATTT 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



QY	6	ThrThrThrAnpHeValAlaGluAsnArgProThrPheGlyGluThrPheAspValMet	25
Db	490	ACGACAGCTTCCCTCAGGGGAGCACCAGGCCCGCTCAGGGATCTCCAC	440
QY	26	ArgGluAlaLeuLeuAsgValLysSerGluArgLeuAlaMetLeuArgAlaLeuAla	45
Db	439	-----ACCATGACTGTGCAATCTCGAGCCACTGGCAATGACGTATCATTTG	392
QY	46	GlyMetCysGlyHisArgValLeuProGlyThrGlyAla	59
Db	391	-----TCGGGGACACAGGCGATGTCACAGCAGAGGGCTGTGTGGCCACACACCGTGGCC	338
QY	60	AlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeuLysProPro	77
Db	337	GCATTTCTTTGTCACACAGCTCCAGTCTTCCCGAGGGCAGCACCAGGAAGCCCTCCCCCG	278
QY	78	ArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMet	97
Db	277	CTGGCTCACATATGAGGGCCACGACTTGGGTTGACAGCACAG	233
QY	98	AsnLysThrIleSerGlnGlu-SerAlaArgValAsnHisArgLeuProGluGlyHisPr	117
Db	232	---AAGCCACTGTCCCAAGTGGTCTGCGAGACCGCAGCATCTCTAACAACACTGGTTCGCC	176
QY	117	oLeuLeu 119	
Db	175	TTGGCTG 169	
RESULT 14			
LOCUS	BG180114/c		
DEFINITION	6032329694F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:443119 5',	607 bp	linear EST 06-FEB-2001
ACCESSION	BG180114		mRNA sequence.
VERSION	BG180114.1	GI:12686817	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 607)		
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished		
	Contact: Robert Strausberg, Ph.D.		
	Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>		
	Tissue Procurement: DCTD/DTP		
	cDNA Library Preparation: Life Technologies, Inc.		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	Plate: LLAM10185 row: k column: 24		
	High quality sequence stop: 555.		
FEATURES	Location/Qualifiers		
source	1..607		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clones="IMAGE:4431119"		
	/tissue_type="adenocarcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH_MGC_91"		
	/note="Organ: prostate; Vector: pCMV-SPORT6; Site:1: NotI;		
	Site:2: SalI; Cloned unidirectionally; oligo-dT primed.		
	Average insert size 1.4 Kb. Library enriched for		
	full-length clones and constructed by Life Technologies.		
	Note: this is a NIH MGC Library."		
BASE COUNT	102 a 188 c 189 g 128 t		
ORIGIN			





High quality sequence stop: 515.

FEATURES		REFERENCE	
Source		AUTHORS	
1. 1172		1. (bases 1 to 660)	
/organism="Homo sapiens"		Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,	
/mol_type="mRNA"		Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamakawa, I., Kiyosawa, H.	
/db_xref="taxon:9606"		Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,	
/clone="IMAGE:5434084"		Gojbori, T., Baidarelli, R., Hill, D.P., Bult, C., Hume, D.A.,	
/tissue_type="lymphoma, cell line"		Quackenbush, J., Schreli, L.M., Kanapin, A., Matsuda, H., Batalov, S.,	
/lab_host="DH10B (phage-resistant)"		Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani	
/clone_lib="NIH_MGC_99"		L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest	
/note="Organ: lymph. Vector: pOTB7; Site: 1: XhoI; Site 2:		A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.	
ecori; cDNA made by oligo-dT priming. Directionally cloned		Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,	
into ecorI/XhoI sites using the following 5' adaptor:		Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M.,	
GGCAGCAG(G). Size-selected >500bp for average insert size		King, B.L., Konegaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons	
1.8kb. Library constructed by Ling Hong in the laboratory		P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki	
of Gerald M. Rubin (University of California, Berkeley)		H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perle, G.,	
using ZAP-cDNA synthesis kit (Stratagene) and Superscript		Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D.,	
II RT (Life Technologies). Note: this is a NIH_MGC		Ranachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring	
Library."		B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Secou	
Library." 169 a 463 c 260 g 280 t		M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale	
BASE COUNT		Watanabe, Y., Wells, C., Wilming, L.G., Wyrshaw-Boris, A., Yanagisawa	
ORIGIN		M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,	
Alignment Scores:		Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura	
Pred. No.:	43.1	Length:	1172
Score:	94.00	Matches:	41
Percent Similarity:	41.79%	Conservative:	15
Best Local Similarity:	30.60%	Mismatches:	43
Query Match:	13.28%	Indels:	35
DB:	12	Gaps:	6
US-10-087-573-2 (1-141) x BM909493 (1-1172)		Contact: Yoshihide Hayashizaki	
QY 4 ThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPheAsp 23		Laboratory for Genome Exploration Research Group, RIKEN Genomic	
Db 513 ACCAATGCCCACAGCTGGTGGCCCTCAGGGTGACGAGGGGCTCCCGAGGGGC 454		Sciences Center (GSC), Yokohama Institute	
QY 24 ValMetArg-----GluAlaLeuArgValValSerSerSerGluArgLeu 38		The Institute of Physical and Chemical Research (RIKEN)	
Db 453 AGCATCAGGCCCCCATCCGCGATCTCCACACCATGCTGTGCGAGTCTCGGAGCCACTG 394		1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan	
QY 39 AlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla 58		Tel: 81-45-503-9222	
Db 393 GCAATGACGTGTGTCATG-----TGGGGCCACCGGCGGCTGTCTGACGAGGGGT 343		Fax: 81-45-503-3216	
QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAla 70		Email: genome-res@gsc.riken.go.jp,	
Db 342 GTGTGGCCACAGACCGTGGGCGCATTTGTCCACACAGTCTTGCACGGGGGCAGC 283		URL: http://genome.gsc.riken.go.jp/	
QY 71 SerMetLysLeuLysProProArgProGlnSerThrLysSerProGluLeuArgGluLeu 90		Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda	
Db 282 ACCAGGAAGCCCTCCCGG-----LysThrIleSerGlnGlu 104		S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,	
QY 91 SerArgLysIleArgGluMetAsn-----LysThrIleSerGlnGlu 104		Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno	
Db 258 GCCTCAGATCAGGCGCCACAACTTAGGTTGACAGCAGACAGGACCTGTCACAGGTG 199		H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.	
QY 105 -SerAlaArgValAsnHisArgLeuProGluGlyHisPro 117		Nunazaki, R., Ohno, M., Ohmoto, N., Saito, R., Sakazume, N., Sano, H.,	
Db 198 GTCTGTGAGCGGCGCATCTTCATGACGTGTCGGCCT 159		Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,	
RESULT 18		Submission	
BY749167		Computational Analysis of Full-Length Mouse cDNAs Compared with	
LOCUS		Human Genome Sequences Mamm. Genome 12, 673-677 (2001)	
DEFINITION		Normalization and subtraction of cap-trapper-selected cDNAs to	
BY749167		prepare full-length cDNA libraries for rapid discovery of new	
GENE		genes. Genome Res. 10 (10), 1617-1630 (2000)	
KEYWORDS		RIKEN integrated sequence analysis (RISA) system--384-format	
SOURCE		sequencing pipeline with 384 multicapillary sequencer. Genome Res.	
ORGANISM		10 (11), 1757-1771 (2000)	
Mus musculus (house mouse)		Computer-based methods for the mouse full-length cDNA	
Mus musculus		encyclopedia: real-time sequence clustering for construction of a	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)	
		cDNA library was prepared and sequenced in Mouse Genome	
		Encyclopedia Project of Genome Exploration Research Group in Riken	
		Genomic Sciences Center and Genome Science Laboratory in RIKEN.	
		Division of Experimental Animal Research in Riken contributed to	
		prepare mouse tissues.	
		Tissues were provided by Dr. John Todd (Dept. of Medical Genetics	
		Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome	
		Trust/MRC Building Addenbrookes Hospital Cambridge) whose	
		assistance we gratefully acknowledge.	
		Please visit our web site ( <a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a> ) for	
		further details.	
		Location/Qualifiers	
		1. .660	

```
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="F630230F14"
/cell_type="NOD-derived CD11c +ve dendritic cells"
/clone_lib="RIKEN full-length enriched, NOD-derived CD11c
+ve dendritic cells"
BASE COUNT      118 a  194 c  198 g  148 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:      23.2      Length:      660
Score:          93.50      Matches:      30
Percent Similarity: 47.73%      Conservative: 12
Best Local Similarity: 34.09%      Mismatches:  17
Query Match:    13.21%      Indels:      29
DB:              14      Gaps:         5

US-10-087-573-2 (1-141) x BY749167 (1-660)

QY 30 LeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGly 49
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 444 ATACTGTGCGAGTCTCTAGAGCCACTGCGCATGACATTGTCATTG-----TGTGGA 394

QY 50 HisArgValLeuProGlyThrGlyAla-----SerAlaIleAlaAla 63
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 393 CACCAGCGGATGCTTAGACACAGGGGCAGTGTGGCGCAGACCGAGGGGCACGTTCTGTCT 334

QY 64 Thr-----ValThrProLysGlyAlaSerMetLysLeuLysProProArgProGlnSer 81
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 333 ACTGTCAGTCTTGCCTAGGGTAGCACCAGGAGGCCCTCCCCCA-----286

QY 82 ThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn-----98
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 285 -----CTGCCCTCAGAGTCAGAGCCATCACTTGGGGTTG 250

QY 99 -----LysThrIleSerGln 103
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 ACAGCGCAGAGAACTGTGCCAA 226

RESULT 19
AV667883
LOCUS
DEFINITION
AV667883 Bos taurus ovary fetus Bos taurus CDNA clone ELOV018D12
5', mRNA sequence.
ACCESSION
AV667883
VERSION
AV667883.1 GI:9932629
KEYWORDS
EST.
SOURCE
Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
TAKASUGA,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
1 (bases 1 to 554)
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
21570554
PUBMED
1171328
COMMENT
Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@occcoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
Location/Qualifiers
```

```
source
1. .554
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="EIOV018D12"
/tissue_type="ovary"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus ovary fetus"
/notes="Vector: pZLi; Site 1: SalI; Site 2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT      165 a  135 c  149 g  105 t
ORIGIN

Alignment Scores:
Pred. No.:      23.2      Length:      554
Score:          92.50      Matches:      35
Percent Similarity: 49.59%      Conservative: 25
Best Local Similarity: 28.93%      Mismatches:  45
Query Match:    13.06%      Indels:      16
DB:              9      Gaps:         5

US-10-087-573-2 (1-141) x AV667883 (1-554)

QY 26 ArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAla 45
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 40 CGGATCTCTCCTCGGTTGAGAACGCGATGATAATTGCTTCTGTGGGATTTGGAG 99

QY 46 GlyMetCysGlyHisArgValLeuProGlyThrGlyAlaSerAlaIleAlaAlaThrVa 65
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 100 -----TGTGGATAAACAAGAAATCCCTGGGCAATCGGTAATCCGCTCCGCGGAGTGAA 153

QY 65 lThrProLysGlyAlaSerMetLysLeuLysProProArgPro-----G1 80
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 154 GAGGAGCGATAAAGAAATGCCGTTGAAGCCACAGACCGCTCCTGCTACAGAGCAG 213

QY 80 nSerThrLysSerProGluLeuArgGlu---LeuSerArgLysIleArgGluMetAsnly 99
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 214 GAGTTGCCACAGCCCGCAGGCTGAGACAGGCTTGGAAACAGAATCTCATAGTATGAATCA 273

QY 99 sThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLe 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 274 GTCCAGAGCTTGAGGAC---AGGATTTACAGGCAACCA-----313

QY 119 uGluLysArgAlaGluTyPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnAr 139
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 314 -CACAAAGAGCTCAGCTGGCAGCAGCAGCTGAAATCGATGAAGAACCACTCAGTAAAGCA 372

QY 139 g 139
Db 373 A 373

RESULT 20
AV597776
LOCUS
DEFINITION
AV597776 Bos taurus cartilage fetus Bos taurus CDNA clone
E1C034A05 5', mRNA sequence.
ACCESSION
AV597776
VERSION
AV597776.1 GI:9715261
KEYWORDS
EST.
SOURCE
Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
TAKASUGA,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
1 (bases 1 to 578)
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
21570554
PUBMED
1171328
COMMENT
Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@occcoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
Location/Qualifiers
```

```

PUBMED
COMMENT
11713328
Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazuugi@cocoa.ocn.ne.jp
Single pass sequencing
This clone was obtained from a polyA-deleted cDNA library.
FEATURES
source
1..578
Location/Qualifiers
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1CA034A05"
/tissue_type="cartilage"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus cartilage fetus"
/notes="Vector: pZ11; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 163 a 147 c 173 g 95 t
ORIGIN
Alignment Scores:
Pred. No.: 24.5 Length: 578
Score: 92.50 Matches: 35
Percent Similarity: 49.59% Conservative: 25
Best Local Similarity: 28.93% Mismatches: 45
Query Match: 13.06% Indels: 16
DB: 5
Gaps: 5
US-10-087-573-2 (1-141) x AV597776 (1-578)
QY 26 ArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAla 45
Db 144 CGGATCCTCTCCCTCGGTTGGAACGCGAGTAAATGCTTCGTGCGGATTGGAG 203
QY 46 GlyMetCysGlyHisArgValLeuPro-GlyThrGlyAlaSerAlaIleAlaAlaThrVa 65
Db 204 -----TGTTGAATAAACAATCCCTGGGCACCTGGGAATCCGCTCTCCGGGAGTGAA 257
QY 65 lThrProLysGlyAlaSerMetLysLeuLysProArgPro-----G1 80
Db 258 GAGGAGCGATAAAAGAAATGCCGCTGAAGCCACAGACACCGTCCCTGTACAGAGCAG 317
QY 80 nSerThrLysSerProGluLeuArgGlu---LeuSerArgLysIleArgGluMetAsnLy 99
Db 318 GAGTTGCCACAGCCCGAGCTGAGACAGGGCTCGAACAGAAATCTGATAGTGAATCA 377
QY 99 sThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLe 119
Db 378 GTCCACAGAGCTTGAGGAAC--AGGATTCTACACAGGCAACCA-----417
QY 119 uGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnAr 139
Db 418 -CACACAAGAGCTCAGCTGGCAGCAGCAGCTGAATCGATGAAGAACCAAGTCAGTAAGCA 476
QY 139 g 139
Db 477 A 477
RESULT 21
AV616040
LOCUS
DEFINITION AV616040 Bos taurus ovary fetus Bos taurus cDNA clone E1OV002H02
5', mRNA sequence.
ACCESSION AV616040
VERSION AV616040.1 GI:9751710
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
REFERENCE
AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
1 (bases 1 to 578)
TITLE poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE 21570554
PUBMED 11713328
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazuugi@cocoa.ocn.ne.jp
Single pass sequencing
This clone was obtained from a polyA-deleted cDNA library.
FEATURES
source
1..578
Location/Qualifiers
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1OV002H02"
/tissue_type="ovary"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus ovary fetus"
/notes="Vector: pZ11; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 168 a 142 c 156 g 111 t
ORIGIN
Alignment Scores:
Pred. No.: 24.5 Length: 578
Score: 92.50 Matches: 35
Percent Similarity: 49.59% Conservative: 25
Best Local Similarity: 28.93% Mismatches: 45
Query Match: 13.06% Indels: 16
DB: 5
Gaps: 5
US-10-087-573-2 (1-141) x AV616040 (1-578)
QY 26 ArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArgAlaLeuAla 45
Db 58 CGGATCCTCTCCCTCGGTTGGAACGCGAGTAAATGCTTCGTGCGGATTGGAG 117
QY 46 GlyMetCysGlyHisArgValLeuPro-GlyThrGlyAlaSerAlaIleAlaAlaThrVa 65
Db 118 -----TGTTGAATAAACAATCCCTGGGCACCTGGGAATCCGCTCTCCGGGAGTGAA 171
QY 65 lThrProLysGlyAlaSerMetLysLeuLysProArgPro-----G1 80
Db 172 GAGGAGCGATAAAAGAAATGCCGCTGAAGCCACAGACACCGTCCCTGTACAGAGCAG 231
QY 80 nSerThrLysSerProGluLeuArgGlu---LeuSerArgLysIleArgGluMetAsnLy 99
Db 232 GAGTTGCCACAGCCCGAGCTGAGACAGGGCTCGAACAGAAATCTGATAGTGAATCA 291
QY 99 sThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLe 119
Db 292 GTCCACAGAGCTTGAGGAAC--AGGATTCTACACAGGCAACCA-----331
QY 119 uGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnAr 139
Db 332 -CACACAAGAGCTCAGCTGGCAGCAGCAGCTGAATCGATGAAGAACCAAGTCAGTAAGCA 390
QY 139 g 139
Db 391 A 391

```

```
RESULT 22
AV601755      594 bp  mRNA  linear  EST 27-NOV-2001
LOCUS        AV601755 Bos taurus kidney fetus Bos taurus cDNA clone EIKI006B10
DEFINITION   5', mRNA sequence.
ACCESSION    AV601755
VERSION      AV601755.1 GI:9724077
KEYWORDS     EST.
SOURCE       Bos taurus (cow)
ORGANISM     Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 594)
Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
JOURNAL      21570554
MEDLINE      11713328
PUBMED
COMMENT      Contact: Yoshikazu Sugimoto
              Animal Genetics Division
              Shitakawa Institute of Animal Genetics
              Odakura Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
              Tel: 81-248-25-5641
              Fax: 81-248-25-5725
              Email: kazusugi@cocoa.ocn.ne.jp
              Single pass sequencing.
              This clone was obtained from a polyA-deleted cDNA library.
              Location/Qualifiers
                1..594
                  /organism="Bos taurus"
                  /mol_type="mRNA"
                  /db_xref="taxon:9913"
                  /clone="EIKI006B10"
                  /tissue type="kidney"
                  /dev stage="fetus"
                  /lab_host="DH10B"
                  /clone_lib="Bos taurus kidney fetus"
                  /notes="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
                    was deleted from a NotI site"
BASE COUNT   170 a 152 c 160 g 112 t
ORIGIN
source
Alignment Scores:
Pred. No.:    25.4      Length:    594
Score:        92.50     Matches:    35
Percent Similarity: 49.59%      Conservative: 25
Best Local Similarity: 28.93%      Mismatches: 45
Query Match:   13.06%      Indels:    16
DB:            9        Gaps:      5

US-10-087-573-2 (1-141) x AV601755 (1-594)

QY 26 ArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaVetLeuArgAlaLeuAla 45
|||||
73 CGGGATCCCTCTCCCTCGGGTGGAGACGAGTGAATGCTCTGTGTGGGATTGGAG 132
|||||
46 GlyMetCysGlyHisArgValLeuPro-GlyThrGlyAlaSerAlaAlaAlaAlaThrVa 65
|||||
133 -----TGTTGGATAAACAAGATCCCTGGGCACCTGGGAATCGCTCTCCGGGGAGTGA 186
|||||
65 lThrProLysGlyAlaSerMetLysLeuLysProProArgPro-----G1 80
|||||
187 GAGGAGCGATAAAGAAATGCCCGTGAAGCCACAGACACCGTCCCTCTACAGAGCAG 246
|||||
80 nSerThrLysSerProGluLeuArgGlu---LeuSerArgLysIleArgGluMetAsnly 99
|||||
247 GAGTTCACACAGCCCGAGCTGAGACAGGGTCTGGACAGAGATCTGATAGTAGTGAATCA 306
|||||
```

```
QY 99 sThrIleSerGlnGluSerAlaArgValAsnHisArgLeuProGluGlyHisProLeuLe 119
|||||
307 GTCCAGAGCTTGAGGAAC---AGATTCTACACAGCAACCA----- 346
|||||
QY 119 uGluLysArgAlaGluTyrPheArgHisLeuArgSerLeuLysSerGlnGlyValAsnAr 139
|||||
347 -CACAAACAGCTCAGCTGGCAGCAGCAGCTGAATCGATGAAGAAGACCAAGCAGTCA 405
|||||
139 g 139
406 A 406

RESULT 23
BF371432
LOCUS        BF371432
DEFINITION   RCO-FN0140-200600-021-e01 FN0140 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BF371432
VERSION      BF371432.1 GI:11333457
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 615)
Dias Neto,E., Garcia Correa,R., Vertovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., Geoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC0&t2=RC0-FN0140-
200600-021-e01&t3=2000-06-20&t4=1)
200600-021-e01&t3=2000-06-20&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 36
High quality sequence stop: 560.
Location/Qualifiers
  1..615
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /dev_stage="Adult"
    /clone_lib="FN0140"
    /note="Organ: prostate normal; Vector: puc18; Site 1: SmaI
      ; Site 2: SmaI; A mini-library was made by cloning
      products derived from CRESTES PCR (U.S. Letters Patent
      application No. 196,716 - Ludwig Institute for Cancer
      Research) profiles into the pUC 18 vector. Reverse
      transcription of tissue mRNA and cDNA amplification were
      performed under low stringency conditions."
BASE COUNT   118 a 201 c 186 g 110 t
ORIGIN
Alignment Scores:
Pred. No.:    26.6      Length:    615
Score:        92.50     Matches:    38
Percent Similarity: 46.58%      Conservative: 30
Best Local Similarity: 26.03%      Mismatches: 52
Query Match:   13.06%      Indels:    26
```



DEFINITION	603058924F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5208437 5', mRNA sequence. --
ACCESSION	BI771910
VERSION	BI771910.1 Gi:15763488
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1. (bases 1 to 691)
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

```

clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM1522 row: p column: 06
High quality sequence stop: 689.
Location/Qualifiers
1. .691
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5208437"
/lab_host="DH10B"
/clone_lib="NIH MGC 122"
/notes="Organ: pooled lung and spleen; Vector: pCMV-SPORTS;
Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 036. Note:
this is a NIH MGC Library."
120 a 215 c 219 g 137 t

```

```

BASE COUNT
120 a 215 c 219 g 137 t
/clone lib="NIH_MGC_122"
"note="Organ: pooled lung and spleen; Vector: pCMV-SPORTS;
Site_1: Not1; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dt
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."

```

BASE COUNT 120 a 215 c 219 g 137 t

BASE COUNT		120 a	215 c	219 g	137 t	this is a NIH_MGC Library."	
ORIGIN							
Alignment Scores:							
	Pred. No.:	30.8	Length:	691			
	Scores:	92.50	Matches:	35			
	Percent Similarity:	45.19%	Conservative:	12			
	Best Local Similarity:	33.65%	Mismatches:	27			
	Query Match:	13.06%	Indels:	30			
	DB:	12	Gaps:	5			
US-10-087-573-2 (1-141) x BI771910 (1-691)							

US-10-087-573-2 (1-141) x BI771910 (1-691)

30 LeuArgGValIlySerSerGIuArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGly  
 ...  
 ...  
 ...

Ddb  
396 ATGACTGTGCAGTCCTCGGAGCCACTGGCAATGACGTTGTCAATG-----TGC GGG 346

SECRET

xy 30 hisaigvaileufroglynihiGiyAia-----SerAlaIleAlaAia 63  
|||::: |||::: |||::: |||::: |||:::  
|||::: |||::: |||::: |||::: |||:::

Db 345 CACCAGGCGATGTCTAGCACAGGGGCTGTGTGGCCACAGACCGTGGGCGCATTCCTGTCC 286

15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047

```
87 val i m f l o y s g i v a t e r m e t l y s l e u l y b p r o a r g p r o g i n s e r 81
```

db  
285 ACACGTCCAGTCTTGCCACGGGGCAGCACAGGAGGCCCTCCCCCG----- 238

[illegible]

```

02 IMLYBSEIFROGIBUENWARGGLULenSERARGLysIleArgGLUTMetAsn----- 98
      |||::: ::||| |

```

db  
237 -----CTGGCCTCACAGATCAGGGCCACAAACTTAGGGTTG 202

00

CONFIDENTIAL

Lysine-Arginine-Serine-Alanine-Glutamine-Histidine-Proline

[illegible]

[illegible]



Location/Qualifiers  
1. .1104

```

1. /url="/organs="Tetraodon nigroviridis"
source
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/cclone="011F18"
/cclone_lib="G"
/note="Genoscope sequence ID : COB
220 a 309 C 343 g 230 t 2
ORIGIN

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DBs:
Length:
Matches:
Conservative:
Mismatch:
Indels:
Gaps:
```

US-10-087-573-2 (1-141) x CNS03BBW (1-1104)

```

14 AsnArgProThrPheGlyGluThrPheAspValMetArgGluAlaLeuLeuArgValLys 33
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
463 AGCCGACCTTGATTTTCTGAC-----CTTCGCCTGTGGGCTCAGAGCTCAG 416
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

34 SerSerGluArgLeuAlaMetLeuArgAla-LeuAlaGlyMetCysGlyHisArgValLe 53
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
415 ---ACGAGACGGCTGCTGAGGTCCTCCATCCACGAGGGCCCTCTGCCAGGCT---TG 362

53 uProGlyThrGlyAlaSerIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 72
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
361 GCCTGGGCGCGCTCTACCCGCGTCTCTGAGCGCTCTCATCACCCGAGGTTGTT--- 306
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

361	GCCTGGCGCGGTCTTACCGCCGCTTCTGCAGGCCCTGCTCATCCCGCAGCGTGT	----	306
72	etLysLeuLysProProArgProGlnSerThrLysSerPro	-----	GlulLeuA 88
105	-----CCGCTTCTTCCACAGCTCAGGCTTAATCTACACGCTCACTCA	-----	266
53	uproGlyIntrGlyAlaSerAlaIle---Ala-AlaThrValThrProLysGlyAlaSerM	-----	72
QY			
Db			

88	rgGluLeuSerArgLysIleA	rgGluMetAsnIysThrIle	SerGlnGluserAlaArg-	107
265	GGGATCTGCAAAAAAAC	CACAAAAATCTGGCAGAA	CAACCTCTCAGAAAGAATTAAAG	206
108	-ValAsnHisArgLeuPro	GluGlyHis-----	ProteinLeuGluLys-----	121
205	CTGTGAACCCACCAAC	CTGAGGCTCATCTTCT	CCCCCTGGCAGGAGGACGCTTAG	146
305	CCGGCTCTCTCAGCT	CAGCAGCCATTAATCT	ATCAGAGGTTCAGTCA	250

QY	122	-----ArgAlaGluTyrPheArgHisLeu	129
Ddb	145	GGCCACGCCGAAGGTTCGCCGGTTGAAATCTGCAGCAGCAGCATTTTCCCGTAGCTC	90
RESULT	30		
	BY224833/c		
LOCUS	BY224833	425 bp	mRNA linear EST 10-DEC-2000
DEFINITION	BY224833	RIKEN full-length enriched	pooled tissues, spleen and
		thymus, lymphoblastoid cell line, PBMC,	TCR gene rearrangement

ACCESSION	thymus lymphocytes Mus musculus cDNA clone I6C0006W22 5', mRNA sequence.
VERSION	BY224833
KEYWORDS	BY224833.1 GI:26405942
EST.	
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 425)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nakido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.  
Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,  
Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,  
Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani,  
L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest,  
A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Giasi, C., Godzik, A.,  
Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,

Score:	91.50	Matches:	30
Percent Similarity:	46.59%	Conservative:	11
Best Local Similarity:	34.09%	Mismatches:	18
Query Match:	12.92%	Indels:	29
DB:	13	Gaps:	5
US-10-087-573-2 (1-141) x BY224833 (1-425)			
Qy	30	LeuArgValLysSerSergLueArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGly	49
	:::	:::	:::
Db	421	ATACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTTCATTG	371
	:::	:::	:::
Qy	50	HisArgValLeuProGlyThrGlyVala	63
	:::	:::	:::
Db	370	CACCAAGGCGATGCTAGCACAGGGGCGATGTGGCCGCAGACAGGGGCACGTTCTTTGTCT	311
	:::	:::	:::
Qy	64	Thr-----ValThrProLysGlyAlaSerMetLysLysProArgProGlnSer	81
	:::	:::	:::
Db	310	ACTCGTCAGTCTTGCCTAGGGGTAGCAGCAGGAGCCCTCCCCCA	253
	:::	:::	:::
Qy	82	ThrLysSerProGluLeuArgGluLeuSerArgLysIleArgGluMetAsn	98
	:::	:::	:::
Db	262	-----CTGGCCTCACAGATCAGAGCCATGAACTTGGGGTTG	227
	:::	:::	:::
Qy	99	-----LysThrIleSerGln	103
	:::	:::	:::
Db	226	ACAGCGCCAGAGCCACTGTCCAA	203

RESULT 31	
LOCUS	BE553346/c
DEFINITION	BE553346 544 bp mRNA linear EST 15-AUG-2000 U48P04.Y1 NCI CGAP Mam2 Mus musculus cDNA clone IMAGE:3153199.5; similar to SW_CORO_BOVIN Q92176 CORONIN-LIKE PROTEIN P57.; mRNA sequence.
ACCESSION	BE553346
VERSION	BE553346
KEYWORDS	EST.
SOURCE	BE553346.1 GI:9817833
ORGANISM	Mus musculus (house mouse)
	Mus musculus
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NCBI (bases 1 to 544) NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
REFERENCE	Unpublished
AUTHORS	Contact: Robert Strausberg, Ph.D.
TITLE	Email: cgapbs@mail.nih.gov
JOURNAL	This clone is available royalty-free through LNL; contact the IMAGE Consortium ( <a href="mailto:info@image.lnl.gov">info@image.lnl.gov</a> ) for further information.
COMMENT	

```

Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: c9apbs@email.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1055955
Seq primer: -40RP from Gibco
High quality sequence stop: 363.
Location/Qualifiers
1..544
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:3153199"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam2"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
101 a 154 c 176 q 111 t
BASE COUNT

```

```

/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:315319"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam2"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Öligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT      101 a      154 c      176 g      111 t      2 others
ORIGIN
Alignment Scores:      28.5      Length:      544
Pred. No.:

```

ORIGIN	Alignment Scores:	
	Pred. No.:	Length:
	28.5	544

Score: 91.50 Matches: 35  
 Percent Similarity: 42.61% Conservativity: 14  
 Best Local Similarity: 30.43% Mismatches: 31  
 Query Match: 12.92% Indels: 35  
 DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x BE553346 (1-544)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 DB 357 TCCAGGTGACGACAGGCTCCCGAGGGCAGCAGCCAGCCCATCCGGATCTCCAC 298  
 QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 DB 297 ACCATA-----ACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTG 256  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 DB 255 TCATTG-----TGTGGACACAGCGGAGTCTAGCAGAGGGGAGTGTGGCCGAG 205  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 DB 204 ACCAGGGGCAGGTTCTTGTCTACTCGTCCAGTCTTGTCCAGGTTAGCAGGAGGCC 145  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 DB 144 CTCTCCCA-----CTGGCTCAGATC 121  
 QY 95 ArgGluMetAsn-----LysThrIleSerGln 103  
 DB 120 AGAGCCATGAACCTGGGGTTGACGCGCAGAGCCACTGTCCCA 76

RESULT 32  
 BE284887/c  
 LOCUS 601097563F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3496502 5',  
 DEFINITION mRNA sequence.

ACCESSION BE284887  
 VERSION BE284887.1 GI:9162136  
 KEYWORDS EST.

SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 640)  
 NIH-MGC <http://mgi.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM8548 row: m column: 15  
 High quality sequence stop: 612.  
 Location/Qualifiers

FEATURES  
 source

1..640  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3496502"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stages="7 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP\_Mam5"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigators

providing samples: Lothar Hennighausen/Robin Humphreys,  
 NIH  
 BASE COUNT 114 a 188 c 202 g 136 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 35.1 Length: 640  
 Score: 91.50 Matches: 35  
 Percent Similarity: 42.61% Conservativity: 14  
 Best Local Similarity: 30.43% Mismatches: 31  
 Query Match: 12.92% Indels: 35  
 DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x BE284887 (1-640)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 DB 444 TCCAGGTGACGACAGGCTCCCGAGGGCAGCAGCCAGCCCATCCGGATCTCCAC 385  
 QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 DB 384 ACCATA-----ACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTG 343  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 DB 342 TCATTG-----TGTGGACACAGCGGATCTTAGCAGAGGGGAGTGTGGCCGAG 292  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 DB 291 ACCAGGGGCAGGTTCTTGTCTACTCGTCCAGTCTTGTCCAGGTTAGCAGGAGGCC 232  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 DB 231 CTCTCCCA-----CTGGCTCAGATC 208  
 QY 95 ArgGluMetAsn-----LysThrIleSerGln 103  
 DB 207 AGAGCCATGAACCTGGGGTTGACGCGCAGAGCCACTGTCCCA 163

RESULT 33  
 BE289370/c  
 LOCUS 601089338F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3484185 5',  
 DEFINITION mRNA sequence.

ACCESSION BE289370  
 VERSION BE289370.1 GI:9169696  
 KEYWORDS EST.

SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 650)  
 NIH-MGC <http://mgi.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM8516 row: 1 column: 10  
 High quality sequence stop: 589.  
 Location/Qualifiers

FEATURES  
 source

1..650  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3484185"

/tissue\_type="tumor, gross tissue"  
 /dev\_stage="7 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Mam5"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Robin Humphreys,  
 NIH"

BASE COUNT 107 a 197 c 210 g 136 t

## ALIGNMENT SCORES:

Pred. No.: 35.8 Length: 650  
 Score: 91.50 Matches: 35  
 Percent Similarity: 42.6% Conservative: 14  
 Best Local Similarity: 30.43% Mismatches: 31  
 Query Match: 12.92% Indels: 35  
 DB: 10 Gaps: 6

US-10-087-573-2 (1-141) x BE289370 (1-650)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 |||||  
 DB 457 TCCAGGTGACGACAGCTCCCGCAAGGCACACAGGCCCCCATCCGGATCTCCAC 398  
 |||||  
 QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 |||||  
 DB 397 ACCATA-----ACTGTGCAGTCTCCAGAGCCACTGGCAATGACATTG 356  
 |||||  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 |||||  
 DB 355 TCATTG-----TGTGGACACAGCGCATGCTTAGCACAGGGGCGAGTGTGGCGGCGAG 305  
 |||||  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 |||||  
 DB 304 ACCAGGGGACGCTCTGTCTACTCTCCAGTCTTGCTTAGGCTAGGGTAGCACAGGAAGGCC 245  
 |||||  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLeu 94  
 |||||  
 DB 244 CTCTCCCA-----CTGGCCTCAGATC 221  
 |||||

RESULT 34  
 BI150185/c  
 LOCUS 602848591F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:5011991 5',  
 DEFINITION mRNA sequence.

ACCESSION BI150185  
 VERSION BI150185.1 GI:14610186  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 (bases 1 to 674)  
 NIH-MGC http://mgs.nci.nih.gov/.

AUTHORS  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
 COMMENT Unpublished

Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 cDNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM1062 row: f column: 24  
 High quality sequence start: 2

High quality sequence stop: 674.

FEATURES  
 source  
 1..674  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5011991"  
 /tissue\_type="spontaneous tumor, metastatic to mammary."  
 Stem cell origin.  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP\_Lu29"  
 /notes="Organ: lung; Vector: pCMV-SPORT6; Site 1: Sali;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

BASE COUNT 120 a 192 c 224 g 138 t

## ALIGNMENT SCORES:

Pred. No.: 37.5 Length: 674  
 Score: 91.50 Matches: 35  
 Percent Similarity: 42.61% Conservative: 14  
 Best Local Similarity: 30.43% Mismatches: 31  
 Query Match: 12.92% Indels: 35  
 DB: 12 Gaps: 6

US-10-087-573-2 (1-141) x BI150185 (1-674)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 |||||  
 DB 451 TCCAGGTGACGACAGCTCCCGCAAGGCACACAGGCCCCCATCCGGATCTCCAC 392  
 |||||  
 QY 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 |||||  
 DB 391 ACCATA-----ACTGTGCAGTCTCCAGAGCCACTGGCAATGACATTG 350  
 |||||  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
 |||||  
 DB 349 TCATTG-----TGTGGACACAGCGCATGCTTAGCACAGGGGCGAGTGTGGCGGCGAG 299  
 |||||  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 |||||  
 DB 298 ACCAGGGGACGCTCTGTCTACTCTCCAGTCTTGCTTAGGCTAGGCGTAGCACAGGAAGGCC 239  
 |||||  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysLeu 94  
 |||||  
 DB 238 CTCTCCCA-----CTGGCCTCAGATC 215  
 |||||

## RESULT 35

BI150185/c

LOCUS

DEFINITION

602099446F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4219522 5',  
 mRNA sequence.

ACCESSION BI150185

VERSION BI150185.1 GI:11655612

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 727)  
 NIH-MGC http://mgs.nci.nih.gov/.

AUTHORS

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

COMMENT Unpublished

Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.



REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 786)  
 NIH-MGC http://mgs.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov  
 Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM1165 row: a column: 11  
 High quality sequence stop: 764.

# FEATURES source

1..786  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="129,C57BL/6J,FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5061010"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="10 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP\_Mam3"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."  
 146 a 230 c 245 g 165 t

Alignment Scores:  
 Pred. No.: 45.7 Length: 786  
 Score: 91.50 Matches: 35  
 Percent Similarity: 42.6% Conservative: 14  
 Best Local Similarity: 30.4% Mismatches: 31  
 Query Match: 12.9% Indels: 35  
 DB: 12 Gaps: 6

US-10-087-573-2 (1-141) x BI158106 (1-786)

QY 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
 |||||  
 Db 456 TCCAGGTGACGACAGGCTCCGCGAGGCGACACAGCCCGCCATCCGGATCTCCAC 397  
 |||||  
 QY 23 AspValMetArgGluAlaLeuLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42  
 |||||  
 Db 396 ACCATA-----ACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTG 355  
 |||||  
 QY 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyVala----- 58  
 |||||  
 Db 354 TCATTG-----TGTGGACACAGCGGATGTCTAGCAGGGGGAGTGTGGCCGAG 304  
 |||||  
 QY 59 -----SerAlaIleAlaAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
 |||||  
 Db 303 ACCAGGGGACGCTCTTGCTACTCGTCCAGTCTTGCTAGGGGTAGCAGGAGGCC 244  
 |||||  
 QY 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
 |||||  
 Db 243 CTGCCCCCA-----CTGGCCTCACAGATC 220  
 |||||  
 QY 95 ArgGluMetAsn-----LysThrIleSerGln 103  
 |||||  
 Db 219 ACAGCATGAACCTGGGGTTGACAGCGCAGAGACCACTGTCCCAA 175  
 |||||

RESULT 38  
 BI839165/c

LOCUS  
 DEFINITION

BI839165 835 bp mRNA linear EST 04-OCT-2001  
 603086487F1 NIH\_MGC\_120 Homo sapiens CDNA clone IMAGE:5225661 5',  
 mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

BI839165  
 BI839165.1 GI:15950715  
 EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

1 (bases 1 to 835)  
 NIH-MGC http://mgs.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM1167 row: m column: 22  
 High quality sequence start: 22  
 High quality sequence stop: 587.

# FEATURES source

1..835  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5225661"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_120"  
 /note="Organ: pooled pancreas and spleen; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH\_MGC Library."  
 160 a 245 c 254 g 176 t

# Alignment Scores:

Pred. No.: 49.2 Length: 835  
 Score: 91.50 Matches: 42  
 Percent Similarity: 38.41% Conservative: 16  
 Best Local Similarity: 27.81% Mismatches: 34  
 Query Match: 12.92% Indels: 59  
 DB: 12 Gaps: 9

US-10-087-573-2 (1-141) x BI839165 (1-835)

QY 15 ArgProThrPheGlyGluThrPheAspVal-----MetArgGluAlaLeu--- 29  
 |||||  
 Db 538 AGGCCACCATGCCCCACACGCTTGTGTGACCCCTCCAGGGGTGACAGCGGCTCCCG 479  
 |||||  
 QY 30 -----LeuArgValLysSerSer 35  
 |||||  
 Db 478 CAGGGGAGCATCAGCCCCCATCCCGGATCTCCACACCATGCTGTGAGTCTCTCG 419  
 |||||  
 QY 36 GluArgLeuAlaMetLeuArgAlaLeuAlaGlyMetCysGlyHisArgValLeuProGly 55  
 |||||  
 Db 418 GAGCCATGGCAATGACGTTGTGATTG-----TGCGGGCACCAGCGGATGTCTAGC 368  
 |||||  
 QY 56 ThrGlyVala-----SerAlaIleAlaAlaThr-----ValThrPro 67  
 |||||  
 Db 367 ACAGGGCTGTGTGGCCACAGACCGTGGGGCGCATTTGTTCACAGTCCAGTCTTGCCC 308  
 |||||  
 QY 68 LysGlyAlaSerMetLysLeuLysProArgProGlnSerThrLysSerProGluLeu 87

```
Db 307 AGGGCAGCACCAAGAGGCCCTCCCGG----- 278
Qy 88 ArgGluLeuSerArgLysIleArgGluMetAsn-----LysThrIle 101
Db 277 -----CTGGCTCACAGATCAGGCCCAAACTTAGGGTTGACAGCACAGGCCACTG 224
Qy 102 SerGlnGluSerAlaArgValAsnHis-----ArgLeu 112
Db 223 TCCCAAGTGGTCTGTGAGACAGCGCACATCTTCATAGCACTGGTGGCGCTTGGCGGCTG 164
Qy 113 ProGlu-----GlyHisProLeuLeuLys 121
Db 163 TCCCAACACGTGGCGGAACCTGCTGGAGCGG 131

RESULT 39
B1411683/c 849 bp mRNA linear EST 14-AUG-2001
LOCUS 602965493F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5120758 5',
DEFINITION mRNA sequence.
ACCESSION B1411683
VERSION B1411683.1 GI:15172606
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 849)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1294 row: j column: 23
High quality sequence start: 3
High quality sequence stop: 794.
Location/Qualifiers
1..849
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5120758"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NCI_CGAP_Lu33"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTTCACCATCTGAGTGGAGCGCGCGCTGTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

```
BASE COUNT 163 a 232 c 272 g 182 t
ORIGIN
Alignment Scores:
Pred. No.: 50.4 Length: 849
Score: 91.50 Matches: 35
Percent Similarity: 42.61% Conservative: 14
Best Local Similarity: 30.43% Mismatches: 31
Query Match: 12.92% Indels: 35
```

```
Db 12 Gaps: 6
US-10-087-573-2 (1-141) x B1411683 (1-849)
Qy 3 SerThrSerThrThrThrAsnPheValAlaGluAsnArgProThrPheGluThrPhe 22
Db 351 TCCAAGGTGACGACGCTCCCGCAAGGCGAGCAGCAGGCCCGCCCTCCGATCTCCAC 292
Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgLeuAlaMetLeuArg 42
Db 291 ACCATA-----ACTGTGCAGTCTCTCAGAGCCACTGGCAATGACATTG 250
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58
Db 249 TCATTG-----TGTGGACACAGCGCGATCTAGCACAGCGCGAGTGTGGCGCGAG 199
Qy 59 -----SerAlaIleAlaIleThr-----ValThrProLysGlyAlaSerMetLysLeu 74
Db 198 ACCAGGGCGACGTTCTTGTCTACTCTGCCAGTCTTCCTAGGGGTAGCAGGAGGCC 139
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94
Db 138 CTCCCCCA-----CTGGCTCACAGATC 115
Qy 95 ArgGluMetAsn-----LysThrIleSerGln 103
Db 114 AGAGCCATGAACCTGGGGTTGACAGCGCGACAGCCACTGTCCCAA 70

RESULT 40
CB586950/c 872 bp mRNA linear EST 03-APR-2003
LOCUS AGENCOURT12989871 NIH_MGC_136 Mus musculus cDNA clone
DEFINITION IMAGE:30292954 5', mRNA sequence.
ACCESSION CB586950
VERSION CB586950.1 GI:29504806
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 872)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM337 row: k column: 11
High quality sequence stop: 661.
Location/Qualifiers
1..872
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30292954"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/notes="Vector: pCMV-Sport6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Normalized, full-length enriched library from pool
of mouse embryonic limb, maxilla and mandible, embryonic
day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and
jaw equivalents from respective days). Cloned
directionally, oligo-dT primed
(5'-GACTAGTCTTATGATCGGAGCGCGCCCTT)15-3'. Size selected
for the 1kb fragments, average insert size 1.2 kb.
Normalization to Cot 7.5. Tissue contributed by David
```

Rowe; library constructed by ResGen, Invitrogen Corp.  
Note: this is a NIH\_MGC Library."

BASE COUNT 165 a 253 c 274 g 180 t  
ORIGIN

Alignment Scores:  
Pred. No.: 52.1 Length: 872  
Score: 91.50 Matches: 35  
Percent Similarity: 42.61% Conservative: 14  
Best Local Similarity: 30.43% Mismatches: 31  
Query Match: 12.92% Indels: 35  
DB: 14 Gaps: 6

US-10-087-573-2 (1-141) x CB586950 (1-872)

Qy 3 SerThrSerThrThrAsnPheValAlaGluAsnArgProThrPheGlyGluThrPhe 22  
Db 443 TCCAAGGTGACGACAGGCTCCCGAAGGCGAGCACCCAGGCCCCCATCCGGGATCTCCAC 384  
Qy 23 AspValMetArgGluAlaLeuArgValLysSerSerGluArgGluAlaMetLeuArg 42  
Db 383 ACCATA-----ACTGTGAGTCTCTAGAGCCACTGGCAATGACATTG 342  
Qy 43 AlaLeuAlaGlyMetCysGlyHisArgValLeuProGlyThrGlyAla----- 58  
Db 341 TCATTG-----TGTGGACACGAGGCGATGCTAGCACAGGGCGCAGTGTGGCGCGAG 291  
Qy 59 -----SerAlaIleAlaThr-----ValThrProLysGlyAlaSerMetLysLeu 74  
Db 290 ACCAGGGGACGTTCTTGTCTACTGTCAGTCTTGCCCTAGGGGTAGCACCAGGAAGGCC 231  
Qy 75 LysProProArgProGlnSerThrLysSerProGluLeuArgGluLeuSerArgLysIle 94  
Db 230 CCTCCCCCA-----CTGGCCTCACAGATC 207  
Qy 95 ArgGluMetAsn-----LysThrIleSerGln 103  
Db 206 AGAGCCATGAACCTGGGGTTGACAGCGCAGAGCCACTGTCCCAA 162

Search completed: November 17, 2003, 14:11:40  
Job time : 1938 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:38:30 ; Search time 41 Seconds  
(without alignments)

545.865 Million cell updates/sec

Title: US-10-087-573-2

Perfect score: 708

Sequence: 1 MESTSTTNFAENRPTGE.....RAEYFRHLRLSKSQVNELI 141

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	708	100.0	141	23 ABP53714	Babesia canis Bcvi
2	638	90.1	285	23 ABP53715	Babesia canis Bcvi
3	83	11.7	192	7 AAP61504	Sequence of p' pro
4	81.5	11.5	945	22 AAU03538	Human protein kina
5	81.5	11.5	945	23 AAU03537	Human kinase polyp
6	81.5	11.5	1259	23 ABG61917	Prostate cancer-as
7	80.5	11.4	1157	22 ABB63351	Drosophila melanog
8	79	11.2	192	20 AAY01973	HIV-1 viral infect
9	79	11.2	1780	22 ABUS3201	Human cell cycle-a

10	79	11.2	1780	22	AA338681	Human polypeptide
11	79	11.2	1788	22	AA40467	Human polypeptide
12	78.5	11.1	828	22	ABG18310	Novel human diagno
13	78.5	11.1	828	22	ABG29103	Novel human diagno
14	78	11.0	203	7	AAP60348	HTLV-III virus (HI
15	77.5	10.9	409	22	AB88134	Drosophila melanog
16	77.5	10.9	451	23	AB883793	Fungal decaprenyl
17	77	10.9	96	20	AAW94135	VIF-derived HIV pr
18	77	10.9	192	20	AAW01976	HIV-1 viral infect
19	77	10.9	192	20	AAW01977	HIV-1 viral infect
20	77	10.9	192	20	AAW01978	HIV-1 viral infect
21	77	10.9	192	20	AAW01970	HIV-1 viral infect
22	77	10.9	192	20	AAW01971	HIV-1 viral infect
23	77	10.9	192	20	AAW01972	HIV-1 viral infect
24	77	10.9	192	20	AAW89324	HIV-1 P protein se
25	77	10.9	192	22	AAW85994	Amino acid sequenc
26	77	10.9	203	7	AAP60421	Sequence of LAV vi
27	77	10.9	203	14	AAW43868	HTLV-III SOR gene
28	77	10.9	203	23	AAO19388	Lymphadenopathy-as
29	76.5	10.8	214	22	AAW3854	Human prostate can
30	76.5	10.8	373	20	AAW50122	Human histone fusi
31	76.5	10.8	373	22	AAW30626	Amino acid sequenc
32	76.5	10.8	1778	22	AAW79480	Human protein SEQ
33	76.5	10.8	2099	22	AAW78496	Human protein SEQ
34	76	10.7	203	14	AAW43876	HTLV-III SOR gene
35	76	10.7	259	22	AAW43752	Human prostate can
36	75.5	10.7	442	22	AB88120	Drosophila melanog
37	75.5	10.7	542	22	ABG19829	Novel human diagno
38	75.5	10.7	1257	24	ABJ19818	Androgen-independe
39	75	10.6	476	22	ABG62927	Drosophila melanog
40	74.5	10.5	1116	22	AB88597	Drosophila melanog
41	74	10.5	192	20	AAW90177	HTLV-III p' protei
42	74	10.5	502	23	AB883792	Fungal decaprenyl
43	73.5	10.4	190	20	AAW01988	Consensus sequence
44	73.5	10.4	191	20	AAW01980	HIV-1 viral infect
45	73.5	10.4	192	20	AAW01979	HIV-1 viral infect

#### ALIGNMENTS

RESULT 1  
ABP53714

ID ABP53714 standard; Protein; 141 AA.

AC ABP53714;

DT 23-DEC-2002 (first entry)

DE Babesia canis Bcvi15 15kd protein SEQ ID NO:2.

XX Babesia canis; Bcvi15; 15kd protein; Bcvi32; 32kd protein; infection;  
XX antiparasitic; immunostimulant; vaccine.

OS Babesia canis.

XX EPI238983-A1.

XX 11-SEP-2002.

PF 04-MAR-2002; 2002EP-0075930.

XX 06-MAR-2001; 2001EP-0200816.

XX (ALKU ) AKZO NOBEL NV.

XX Schettlers TPM, Carcy BPD, Drakulovski PR, Gorenflot AF;

XX WPI; 2002-724917/79.

XX N-PSDB; ABQ82649.

PT Novel Babesia canis associated protein and nucleic acid encoding the  
protein, useful in a vaccine and in the manufacture of vaccines for

```

PT combating Babesia canis infections
XX Claim 1; Fig 2; 41pp; English.
XX
CC The present invention describes a Babesia canis associated protein (I),
CC comprising a BcVir15 protein of 15 kD molecular weight (MW) and having
CC a sequence of at least 80% homology to a sequence (ABP53715) of 141
CC amino acids, or a BcVir32 protein of 32 kD MW and having a sequence of
CC at least 80% homology to a sequence (ABP53715) of 285 amino acids, or
CC their immunogenic fragments. (I) have antiparasitic and immunostimulant
CC activities, and can be used in vaccines. (I) can also be used for the
CC preparation of a vaccine for combating B. canis infections. (I) is also
CC useful in a diagnostic test for the detection of antibodies against
CC B. canis associated antigenic material. The present sequence represents
CC BcVir15 from the present invention.
XX
SQ Sequence 141 AA;
Query Match 100.0%; Score 708; DB 23; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.1e-68;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MESTSTTNFVAENRPTFGTDFVMEALLRVKSSERLALAGMCGHRVLPQTGASA 60
DB 1 MESTSTTNFVAENRPTFGTDFVMEALLRVKSSERLALAGMCGHRVLPQTGASA 60
QY 61 IAATVTPKGASMKLKPDPQSTKSPPELRELSRKIRENMNTTISQESARVNHRLPEGHPLLE 120
DB 61 IAATVTPKGASMKLKPDPQSTKSPPELRELSRKIRENMNTTISQESARVNHRLPEGHPLLE 120
QY 121 KRAEYFRHLRLSKSQVNRLI 141
DB 121 KRAEYFRHLRLSKSQVNRLI 141
RESULT 2
ABP53715
ID ABP53715 standard; Protein; 285 AA.
XX
AC ABP53715;
XX
DT 23-DEC-2002 (first entry)
XX
DE Babesia canis BcVir32 32kD protein SEQ ID NO:4.
XX
KW Babesia canis; BcVir15; 15kD protein; BcVir32; 32kD protein; infection;
XX antiparasitic; immunostimulant; vaccine.
XX
OS Babesia canis.
XX
PN EP1238983-A1.
XX
XX 11-SEP-2002.
XX
PF 04-MAR-2002; 2002EP-0075830.
XX
PR 06-MAR-2001; 2001EP-0200816.
XX
PA (ALKU ) AKZO NOBEL NV.
XX
PI Schettlers TPM, Carcy BPD, Drakulovski PR, Gorenflot AF;
XX
DR N-PSDB; ABQ82650.
XX
XX WPI; 2002-724917/79.
XX
XX Novel Babesia canis associated protein and nucleic acid encoding the
XX protein, useful in a vaccine and in the manufacture of vaccines for
XX combating Babesia canis infections
XX
XX Claim 2; Fig 3; 41pp; English.
XX
XX The present invention describes a Babesia canis associated protein (I),
XX comprising a BcVir15 protein of 15 kD molecular weight (MW) and having

```

---

```

CC a sequence of at least 80% homology to a sequence (ABP53714) of 141
CC amino acids, or a BcVir32 protein of 32 kD MW and having a sequence of
CC at least 80% homology to a sequence (ABP53715) of 285 amino acids, or
CC their immunogenic fragments. (I) have antiparasitic and immunostimulant
CC activities, and can be used in vaccines. (I) can also be used for the
CC preparation of a vaccine for combating B. canis infections. (I) is also
CC useful in a diagnostic test for the detection of antibodies against
CC B. canis associated antigenic material. The present sequence represents
CC BcVir32 from the present invention.
XX
SQ Sequence 285 AA;
Query Match 90.1%; Score 638; DB 23; Length 285;
Best Local Similarity 94.8%; Pred. No. 1e-60;
Matches 128; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 MESTSTTNFVAENRPTFGTDFVMEALLRVKSSERLALAGMCGHRVLPQTGASA 60
DB 1 MESTSTTNFVAENRPTFGTDFVMEALLRVKSSERLALAGMCGHRVLPQTGASA 60
QY 61 IAATVTPKGASMKLKPDPQSTKSPPELRELSRKIRENMNTTISQESARVNHRLPEGHPLLE 120
DB 61 IAATVTPKGASMKLKPDPQSTKSPPELRELSRKIRENMNTTISQESARVNHRLPEGHPLLE 120
QY 121 KRAEYFRHLRLSKSQ 135
DB 121 KRAEYFRHLRLSKSQ 135
RESULT 3
AAP61504
ID AAP61504 standard; Protein; 192 AA.
XX
AC AAP61504;
XX
DT 08-JUN-1991 (first entry)
XX
DE Sequence of p' protein.
XX
KW HIV; LAV; AIDS; diagnosis; vaccine.
XX
OS HTLV-IIIb/H9 cells (ATCC CRL 8543).
XX
PN EPI87041-A.
XX
XX 09-JUL-1986.
XX
PF 23-DEC-1985; 85EP-0309454.
XX
PR 24-DEC-1984; 84US-0685272.
XX
PR 04-DEC-1985; 85US-0805069.
XX
PA (GETH ) GENENTECH INC.
XX
XX Capon DJ, Lasky LA;
XX
XX WPI; 1986-177602/28.
XX
XX N-PSDB; AAN60288.
XX
XX Acquired immune deficiency syndrome polypeptide(s) - obt'd. by
XX molecular cloning etc. and used for diagnosis and in vaccines
XX against virus disease
XX
XX Example; fig 2; 125pp; English.
XX
XX A comparison of AAN60287 with the cDNA of the HTLV-III genome
XX revealed one particular clone, designated p7.11 which contained a
XX DNA sequence encoding this peptide (AAP60308) sequence. This approx.
XX 2.2 kilobase covers the precursor gag region and encodes, 5' to 3',
XX p-12, p-15, p-24 a second p-15 protein and approx. 300 extra base
XX pairs 3' to the gag region (see AAN60288).
XX
XX Sequence 192 AA;

```